THE STOOL MICROBIOTA AND INFANT WHEEZING ILLNESS – THE DRAKENSTEIN CHILD HEALTH STUDY, SOUTH AFRICA

Michelle Rudo Ngwarai
NGWMIC003

SUBMITTED TO THE UNIVERSITY OF CAPE TOWN
In fulfilment of the requirements for the degree
MSc. Medical Microbiology
September 2019

Department of Pathology
Faculty of Health Sciences
UNIVERSITY OF CAPE TOWN

To the Ngwarai's and Kanjere's

You never stopped reminding me that *I can do all things through* ***Christ***.

Supervisors

Dr Mamadou Kaba (MD, MSc, PhD)[1,2]
[1]Division of Medical Microbiology, Department of Pathology, University of Cape Town, Cape Town, South Africa.
[2]Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Cape Town, South Africa.

Dr Lemese Ah Tow (MSc, PhD)[1,2]
[1]Division of Medical Microbiology, Department of Pathology, University of Cape Town, Cape Town, South Africa.
[2]Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Cape Town, South Africa.

Declaration

I, Michelle Ngwarai, hereby declare that the work on which this dissertation/thesis is based on is my original work (except where acknowledgements indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university.

I empower the university to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever.

Signed by candidate

25 September 2019

Acknowledgements

I wish to thank my supervisors Dr. Mamadou Kaba and Dr. Lemese Ah Tow for the opportunity to work on this project. Thank you for your support and supervision throughout my MSc.

I would also like to thank Prof. Heather Zar for providing the opportunity to study samples from the Drakenstein Child Health Study.

My sincerest appreciation to Prof. Sugnet Lubbe for taking the time to assist me with the statistical analysis for this dissertation, right from the beginning. Thank you for your patience and willingness to assist as well as help me understand the statistical analysis section. Thank you to Samson Kilaza from the Computational Biology Group, University of Cape Town, for the bio-informatics support.

To the rest of the school family, Menna, Veronica, Samantha, Shantelle, Rendani and Bosco, thank you so much for your encouragement and companionship. Additionally, I would also like to extend my gratitude to everyone in the Division of Medical Microbiology for being helpful and warm.

I acknowledge the National Research Foundation, South Africa, as well as the Division of Medical Microbiology, University of Cape Town for providing me with funding.

To my loving parents, thank you so much for all your prayers, encouragement and financial support. None of this would have been possible without you. To the rest of my family and friends, thank you for your unconditional love, prayers and inspiration throughout this time.

To my dear husband and best friend, words fail me when I try express how grateful I am for the continued support you have shown me throughout my tertiary education, especially during this Masters. Thank you for the late nights you put in, and proof reading all my work several times.

I am eternally grateful to my Heavenly Father for His love and strength, which He continues to show in my life.

Preface

This dissertation is submitted for the degree of Master of Science in Medicine (MSc Med) in Medical Microbiology at the Division of Medical Microbiology, Department of Pathology, Faculty of Health Sciences, University of Cape Town, South Africa. This study obtained ethical approval from the Human Research Ethics Committee (HREC) of the University of Cape Town, South Africa (748/2015). This study also received financial support from the Bill and Melinda Gates Foundation Global Health Grant (OPP1017641), the National Research Foundation (South Africa) and the Wellcome Trust (United Kingdom). The work reported in this dissertation resulted from a collaborative effort between the Department of Paediatrics and Child Health, Department of Pathology, University of Cape Town, South Africa, and the Department of Statistical Sciences, Stellenbosch University, South Africa.

The aim of this dissertation was to describe the major components of the gut bacterial profile during the first year of life, determine the risk factors influencing that composition and to investigate the association between the stool bacteria and the development of recurrent wheezing in infants. The first chapter introduces the context of this project as well as some literature. The second chapter presents a comprehensive overview of the literature reviewed within the context of this project. The third chapter focuses on the experimental and computational approaches used to generate the sequencing data. The third chapter also focuses on answering the main aims of the dissertation, detailed above. The work described in the third chapter resulted from the following collaborative efforts:

- The MSc candidate performed the nucleic acid extraction and the library preparation of all the faecal samples at the Division of Medical Microbiology, University of Cape Town.
- The sequencing was performed at the Centre for Proteomic and Genomic Research (CPGR, South Africa). Samson Kilaza from the Computational Biology Group, University of Cape Town performed the bioinformatics workflow to clean up the raw sequencing data and to classify the operational taxonomic units.
- The data analysis carried out in this was designed by the MSc candidate and Prof. Sugnet Lubbe from the Department of Statistical Sciences, Stellenbosch University.

The fourth and fifth chapters are the general discussion and general conclusion respectively. These tie in together the different elements of this dissertation. The submitted material is the work of the MSc candidate, unless stated otherwise by acknowledgments.

Table of Contents

List of Figures

List of Tables

Abstract

Background: Wheezing is one of the leading respiratory symptoms in childhood, which with recurrent occurrence can lead to asthma at school age. Few studies, with conflicting findings, have investigated the contribution of stool bacteria in childhood wheezing illness. Additionally, most of the published studies were conducted in high-income countries using small sample sizes and targeted bacterial detection techniques. To address these limitations, we conducted this study to determine the association of the infant stool bacterial diversity and composition with wheezing development.

Methods: We conducted a longitudinal case-control study nested within the Drakenstein Child Health Study, Western Cape, South Africa. We included 140 infants with wheeze (cases) and 140 age-matched controls. Passed faecal samples were collected from infants at birth, whilst aspirated or passed faecal samples were collected at six weeks, six months and 12 months. Deoxyribonucleic acid (DNA) was extracted from all the samples using the manual ZR Fecal DNA MiniPrep™ Kit. Sequencing of the hypervariable V4 region of the 16S rRNA gene was performed using the Illumina MiSeq technique. The resulting sequencing data was subjected to bioinformatic quality control checks and statistical analysis.

Results: One of the main findings from the systematic review of wheezing data was the observed association between the development of atopic wheezing and four bacteria namely *Faecalibacterium, Lachnospira, Veillonella* and *Rothia*. This study included 280 children (n=123 female and n=157 male) recruited from TC Newman (38.5%) and Mbekweni (61.5%) areas. Stool samples were collected from 159 infants at birth, 114 infants at six weeks, 141 infants at six months and 98 infants at 12 months of age. Aspirate and passed stool samples were found to be similar and comparable in terms of the bacterial composition. At age six weeks, 16% (23/140) of the infants included had wheezed once. During the first year, the proportion of infants

with recurrent wheezing (≥ 2 wheezing episodes) was 56% (78/140). Independent factors influencing the composition and diversity of the infants faecal bacteria were feeding practices, age of mothers and living conditions (all $p<0.05$). We did not observe distinct faecal bacterial profiles between infants with and without wheeze at any taxonomic level analysis (all $p>0.05$). In addition, we did not find significant difference between infants with and without wheeze for selected bacteria reported to be associated with wheezing. However, we observed an increased relative abundance of *Proteobacteria* at 12 months of age in infants who had previously wheezed. Additionally, *Lactobacillales* was in apparently significantly higher proportions in recurrent wheezers as compared to infants who wheezed once ($p<0.05$).

Conclusion: This study shows that aspirated stool is a good alternative for passed stool in microbiome studies. Furthermore, it revealed the complex and dynamic nature of the faecal bacteria, as well as factors influencing the faecal bacterial profile during the first year of life. More studies in this cohort, including the use of metagenomic and metabolomic approaches are required to demonstrate the role played by bacterial types and their metabolites in the development of wheezing illness. Finally, because recurrent wheezing is one of the precursors of asthma, there is a need for an additional follow-up of these infants in order to investigate the contribution of the early faecal microbiome in the development of asthma at school age.

Chapter 1 - General Introduction

1.1. Background and rationale

Wheezing illness is one of the major causes of respiratory morbidity and decreased lung function during the first year of life (Panettieri *et al.*, 2008; Fauroux *et al.*, 2014; Moraes *et al.*, 2014). The global prevalence of recurrent wheezing in infants during the first 12 months is about 20% (Nriagu *et al.*, 1999; Sunyer *et al.*, 2001; Smuts *et al.*, 2011). Consequences of early wheeze include hospitalizations, reduced quality of life as well as increased chances of developing recurrent wheezing and asthma later on in life (Kotaniemi-Syrjanen *et al.*, 2003; Arden *et al.*, 2006; Henderson *et al.*, 2008). Recurrent wheezing is not only an early indication of possible childhood asthma, but it also has significant health impacts on young children (Sigurs and Bj, 2000; Acevedo *et al.*, 2012; Ducharme, Tse and Chauhan, 2014). Infant wheezing can be triggered by a variety of factors, some of which include bacterial and viral infections, genetic predisposition and dysbiosis of the lung and gut microbiome (Oddy *et al.*, 2004; Kusel *et al.*, 2007; Yao *et al.*, 2011; Arrieta *et al.*, 2015; Kwong and Bacharier, 2017; Ta *et al.*, 2018).

At present, many research groups have been focusing on the role played by the gut microbiome in the development of infant wheezing (Van Nimwegen *et al.*, 2011; Gosalbes *et al.*, 2013; Arrieta *et al.*, 2015). Previous studies have also observed that the gut microbiome is a key player in the development of the infant immune system (Sharma *et al.*, 2009; Gritz and Bhandari, 2015). These studies also noted that a dysbiosis of the gut microbiome could result in a weaker immune system. Previous epidemiological studies have observed an association between infant wheezing illness and faecal bacteria (Murray, Tannock, Simon, Harmsen, Welling and Custovic, 2005; Penders *et al.*, 2007; Van Nimwegen *et al.*, 2011; Arrieta *et al.*, 2015). Arrieta and colleagues observed that children with recurrent wheezing and at risk of developing asthma had significantly reduced amounts of the genera *Faecalibacterium, Lachnospira, Veillionella* and *Rothia* (Arrieta *et al.*, 2015). Nambu and colleagues reported that if the infants were colonised with *Bacteroides*

at four months of age, they were likely to develop allergy at one year (Nambu, Shintaku and Ohta, 2004). Gut colonisation with *Clostridium difficile* has been associated with a higher risk of developing recurrent wheezing (Penders *et al.*, 2007, 2008; Van Nimwegen *et al.*, 2011).

Nevertheless, the results from previous studies investigating the association between developing wheezing in infancy and the faecal bacteria have been inconsistent. We speculate that these inconsistencies could be due to differences in sample sizes, definitions of recurrent wheezing and experiment analyses carried out (Nambu *et al*., 2004; Murray *et al.*, 2005; Amberbir *et al.*, 2011; Bisgaard *et al.*, 2011; Arrieta *et al.*, 2015). Furthermore, limited studies focusing on the association between faecal bacteria and wheezing illness have been conducted in developing countries. Therefore, large-scale longitudinal birth cohort studies in developing countries, using high-throughput sequencing of the 16S rRNA gene, are needed to provide a community-wide taxonomic assessment of the faecal bacteria and its role in wheezing illness.

1.2. Study hypothesis and aim

The main objective of this Masters project was to investigate the association between the composition of the infant stool bacteria and the development of recurrent wheezing in infants.

1.3. Dissertation outline

This dissertation begins with a systematic review that focuses on the association between gastrointestinal bacteria and the development of infant wheezing. The systematic review included a comprehensive view of the literature available from different regions on this topic up to a particular date. The following chapter then details the experimental flow and the main findings observed in this study. This chapter also reported on the infant bacterial profiles in the first year of life, the risk factors associated with bacterial acquisition and the association between

the infant stool bacteria and wheezing illness. The experimental findings chapter is succeeded by a general discussion and conclusion.

1.4. References

Acevedo, N. *et al.* (2012) 'Particular characteristics of allergic symptoms in tropical environments: follow up to 24 months in the FRAAT birth cohort study', *BMC Pulmonary Medicine*, 12(1), p. 13. doi: 10.1186/1471-2466-12-13.

Amberbir, a. *et al.* (2011) 'Effects of Helicobacter pylori, geohelminth infection and selected commensal bacteria on the risk of allergic disease and sensitization in 3-year-old Ethiopian children', *Clinical and Experimental Allergy*, 41(10), pp. 1422–1430. doi: 10.1111/j.1365-2222.2011.03831.x.

Arden, K. E. *et al.* (2006) 'Frequent detection of human rhinoviruses, paramyxoviruses, coronaviruses, and bocavirus during acute respiratory tract infections.', *Journal of medical virology*, 78(9), pp. 1232–1240. doi: 10.1002/jmv.20689.

Arrieta, M. *et al.* (2015) 'Early infancy microbial and metabolic alterations affect risk of childhood asthma', *Science Translational Medicine*, 7(307), p. 307ra152-307ra152. doi: 10.1126/scitranslmed.aab2271.

Bisgaard, H. *et al.* (2011) 'Reduced diversity of the intestinal microbiota during infancy is associated with increased risk of allergic disease at school age', *Journal of Allergy and Clinical Immunology*, 128. doi: 10.1016/j.jaci.2011.04.060.

Ducharme, F. M., Tse, S. M. and Chauhan, B. (2014) 'Diagnosis, management, and prognosis of preschool wheeze', *The Lancet*, 383(9928), pp. 1593–1604. doi: 10.1016/S0140-6736(14)60615-2.

Fauroux, B. *et al.* (2014) 'Respiratory morbidity of preterm infants of less than 33 weeks gestation without bronchopulmonary dysplasia: a 12-month follow-up of the CASTOR study cohort.', *Epidemiology and infection*, 142(7), pp. 1362–74. doi: 10.1017/S0950268813001738.

Gosalbes, M. J. *et al.* (2013) 'Meconium microbiota types dominated by lactic acid or enteric bacteria are differentially associated with maternal eczema and respiratory problems in infants', *Clinical & Experimental Allergy*, 43(2), pp. 198–211. doi: 10.1111/cea.12063.

Gritz, E. C. and Bhandari, V. (2015) 'The human neonatal gut microbiome: A brief review', *Frontiers in Pediatrics*, 3(March). doi: 10.3389/fped.2015.00017.

Henderson, J. *et al.* (2008) 'Associations of wheezing phenotypes in the first 6 years of life with atopy, lung function and airway responsiveness in mid-childhood', *Thorax*, 63(11), pp. 974–980. doi: 10.1136/thx.2007.093187.

Kotaniemi-Syrjanen, A. *et al.* (2003) 'Rhinovirus-induced wheezing in infancy--the first sign of childhood asthma?', *The Journal of allergy and clinical immunology*, 111(1), pp. 66–71.

Kusel, M. M. H. *et al.* (2007) 'Early-life respiratory viral infections, atopic sensitization, and risk of subsequent development of persistent asthma', *Journal of Allergy and Clinical Immunology*, 119, pp. 1105–1110. doi: 10.1016/j.jaci.2006.12.669.

Kwong, C. G. and Bacharier, L. B. (2017) 'Microbes and the role of antibiotic treatment for wheezy lower respiratory tract illnesses in preschool children.', *Current allergy and asthma reports*, 17(5), p. 34. doi: 10.1007/s11882-017-0701-6.

Moraes, L. S. L. *et al.* (2014) 'Prevalence and clinical characteristics of wheezing in children in the first year of life, living in Cuiabá, Mato Grosso, Brazil', *Revista paulista de pediatria : orgão oficial da Sociedade de Pediatria de São Paulo*, 32(4), pp. 313–319. doi: 10.1016/j.rpped.2014.06.004.

Murray, C. S., Tannock, G. W., Simon, M. A., Harmsen, H. J. M., Welling, G. W., Custovic, A., *et al.* (2005) 'Fecal microbiota in sensitized wheezy and non-sensitized non-wheezy children: a nested case–control study', *Clinical & Experimental Allergy*, 35(6), pp. 741–745.

Nambu, M., Shintaku, N. and Ohta, S. (2004) 'Intestinal microflora at 4 months of age and the development of allergy', *Allergology International*, 53(2), pp. 121–126. doi: 10.1111/j.1440-1592.2004.00315.x.

Van Nimwegen, F. a. *et al.* (2011) 'Mode and place of delivery, gastrointestinal microbiota, and their influence on asthma and atopy', *Journal of Allergy and Clinical Immunology*, 128(5), pp. 948–955. doi: 10.1016/j.jaci.2011.07.027.

Nriagu, J. *et al.* (1999) 'Prevalence of asthma and respiratory symptoms in south-central Durban, South Africa.', *European journal of epidemiology*, 15(8), pp. 747–755. doi: 10.1023/A:1007653709188.

Oddy, W. H. *et al.* (2004) 'The relation of breastfeeding and body mass index to asthma and atopy in children: A prospective cohort study to age 6 years', *American Journal of Public Health*, 94(9), pp. 1531–1537. doi: 10.2105/AJPH.94.9.1531.

Panettieri, R. A. *et al.* (2008) 'Natural history of asthma: Persistence versus progression-does the beginning predict the end?', *Journal of Allergy and Clinical Immunology*, 121(3), pp. 607–613. doi: 10.1016/j.jaci.2008.01.006.

Penders, J. *et al.* (2007) 'Gut microbiota composition and development of atopic manifestations in infancy: the KOALA Birth Cohort Study.', *Gut*, 56(5), pp. 661–667. doi: 10.1136/gut.2006.100164.

Penders, J. *et al.* (2008) 'Toxigenic and non-toxigenic Clostridium difficile : determinants of intestinal colonisation and role in childhood atopic manifestations', *Gut*, 57(7), pp. 1025–1026. doi: 10.1136/gut.2007.143214.

Sharma, R. *et al.* (2009) 'Intestinal microbiota: does it play a role in diseases of the neonate?', *NeoReviews*, 10(4), pp. e166–e179. doi: 10.1542/neo.10-4-e166.

Sigurs, N. and Bj, R. (2000) 'Respiratory syncytial virus bronchiolitis in infancy is an important risk fact or for asthma and allergy at age 7', 161, pp. 1501–1507.

Smuts, H. E., Workman, L. J. and Zar, H. J. (2011) 'Human rhinovirus infection in young African children with acute wheezing.', *BMC infectious diseases*, 11(1), p. 65. doi: 10.1186/1471-2334-11-65.

Sunyer, J. *et al.* (2001) 'Prenatal risk factors of wheezing at the age of four years in Tanzania.', *Thorax*, 56(4), pp. 290–295. doi: 10.1136/thorax.56.4.290.

Ta, L. D. H. *et al.* (2018) 'Establishment of the nasal microbiota in the first 18 months of life: Correlation with early-onset rhinitis and wheezing.', *The Journal of allergy and clinical immunology*, 142(1), pp. 86–95. doi: 10.1016/j.jaci.2018.01.032.

Vael, C. *et al.* (2008) 'Early intestinal Bacteroides fragilis colonisation and development of asthma', *BMC Pulmonary Medicine*, 8, p. 19. doi: 10.1186/1471-2466-8-19.

Yao, M.-M. *et al.* (2011) '[Etiology and risk factors of infantile wheezing].', *Zhongguo dang dai er ke za zhi = Chinese journal of contemporary pediatrics*, 13(3), pp. 195–198.

Chapter 2 – A Systematic Review of the Association of Gastrointestinal Bacteria with the Development of Wheeze in Young Children

2.1. Abstract

Background: There is no comprehensive assessment of published articles on the contribution of faecal bacteria in the development of childhood wheezing illness. This systematic review aimed to address this issue.

Methods: A set of keywords were used to search in four electronic databases (Medline, Scopus, EBSCOhost, and Web of Science) for studies reporting on faecal bacteria and wheezing. Additional articles missed by our search terms strategy were obtained by reading references cited in the eligible studies. The search and selection of eligible studies was conducted by two independent authors. The last literature search was carried out on 19 August 2017.

Results: We included 9 studies conducted in children under five years old, all reporting on wheezing, but with 78% (7/9) of them assessing for recurrent wheezing as an outcome. Most of the studies (89%; 8/9) were cross-sectional, and the median sample size of the included studies was 319 participants (range: 18 - 957). Most of the included studies used culture-independent techniques (67%, 6/9) to characterize the faecal bacteria, and massively parallel 16S rRNA gene sequencing was performed in only two of these studies. The majority of the eligible studies (67%, 6/9) were conducted in Europe [Belgium (n= 1), Netherlands (n= 3), Spain (n= 1), and United Kingdom (n= 1)]. The remaining three studies were conducted in North America [Canada (n= 1)], Asia [Japan (n= 1)] and Africa [Ethiopia (n= 1)]. The detection of *Bifidobacteria* spp. (n= 1 study; OR = 3.65; 95% CI: 1.21 to 11.0) was significantly associated with an increased risk for wheezing. The detection of *Clostridium difficile* (n= 3 studies) was significantly associated with wheezing. In contrast, a single study reported a decreased risk for wheezing (OR = 0.53; 95 % CI 0.30 to 0.92) associated with the genus *Clostridium*. The protective role of FLVR bacteria (*Faecalibacterium, Lachnospira, Veillonella* and *Rothia*) was reported in Canadian children with atopic wheeze at the age 1 year.

Conclusion: Very few studies based on small sample sizes reported significant associations between wheezing illness and faecal bacteria. These findings are of great interest because recurrent wheezing is an important predictor of asthma later in life. Therefore, larger longitudinal birth cohort studies are warranted to study the contribution of faecal bacteria in recurrent wheezing. In addition, these studies should adopt high-throughput sequencing techniques (16S rRNA gene and/or whole genome shotgun sequencing) or high-throughput bacterial culture systems (culturomics) as methods of choice, to determine the faecal bacterial profiles of wheezing children.

2.2. Introduction

Wheezing is defined as a high-pitched whistling sound made while breathing, that usually affects children under the age of five years (Martinez *et al.*, 1995; Lau *et al.*, 2003; Taussig *et al.*, 2003). It is diagnosed based on history of wheeze reported by parents/guardians and/or symptoms evidenced by the physicians (Nambu, Shintaku and Ohta, 2004; Murray, Tannock, Simon, Harmsen, Welling and Custovic, 2005a; Van Nimwegen *et al.*, 2011). The definitions of recurrent wheezing are inconsistent between researchers (Murray, Tannock, Simon, Harmsen, Welling and Custovic, 2005b; Penders, Thijs, P. a van den Brandt, *et al.*, 2007; Arrieta *et al.*, 2015), but it is usually defined as three or more wheezing episodes within 12 months (Murray, Tannock, Simon, Harmsen, Welling, Custovic, et al. 2005; John Penders, Thijs, van den Brandt, Kummeling, Snijders, Stelma, Adams, van Ree, Stobberingh, et al. 2007).

Wheezing episodes in children are often caused by viral and bacterial infections (Ferguson, Whitelaw and Brown, 1992; Martinez *et al.*, 1995; Johnston *et al.*, 1996; Oddy *et al.*, 2004; Arden *et al.*, 2006; Kusel *et al.*, 2007; De Schutter *et al.*, 2012; Jackson *et al.*, 2012). Wheezing typically starts in the first year of life and significantly affects the quality of life of both concerned

infants and their parents/legal guardians (Ungar and Coyte, 2001; Stevens *et al.*, 2003). In addition, it is usually accompanied with multiple pediatric health care facility visits, some of which can lead to hospitalizations (Kotaniemi-Syrjanen *et al.*, 2003; Arden *et al.*, 2006; Henderson *et al.*, 2008). It is estimated that one in every three infants has wheezed at some point in their childhood (Bisgaard and Szefler, 2007). Furthermore, infants who wheezed early in life are more likely to develop recurrent wheezing, one of the risk factors of asthma later in life (Ly *et al.*, 2006; Sigurs *et al.*, 2010; Neuman *et al.*, 2014).

There are numerous risk factors reported in literature that have been associated with the development of wheezing in infancy. Some of the risk factors as determined by meta-analyses are preterm birth, low birth weight, antibiotic exposure in pregnancy and rhinovirus-induced wheezing (Been *et al.*, 2014; Mebrahtu *et al.*, 2015; Zhao *et al.*, 2015; Liu *et al.*, 2017). Meta-analyses and reviews on the role of breastfeeding in infant wheezing have concluded that breastfed infants have lower incidences of wheezing than formula fed infants (Friedman and Zeiger, 2005; Ip *et al.*, 2007; Lodge *et al.*, 2015). Air pollution, which includes maternal smoking and traffic related smoke have also been shown to be associated with wheezing illness in infants (Esposito *et al.*, 2014; Vanker, Gie and Zar, 2018). In addition, the human (lung and gut) microbiome has been associated with the development of childhood wheezing illness (Arrieta *et al.*, 2015; Kwong and Bacharier, 2017; Rosas-Salazar *et al.*, 2018; Ta *et al.*, 2018). Exposure to a specific bacterium or group of bacteria has been associated with either a protective (Arrieta *et al.*, 2015) or non-protective role (Verhulst *et al.*, 2008) of gastrointestinal microbes in wheezing development. The conflicting reports on the contribution of certain faecal bacterial species in the development of childhood wheezing illness and the lack of a comprehensive literature review on this topic led us to write this review. This systematic review aimed to assess the association between the gastrointestinal bacteria and the development of wheezing illness in young children.

2.3. Methodology

2.3.1. Literature search strategy

We systematically searched through four electronic databases (Medline, Scopus, EBSCOhost, and Web of Science) for peer-reviewed articles reporting on the detection of bacteria in faeces of infants with wheezing (Table 2.1). We assessed for eligibility in all articles published in English up until 19 August 2017. To identify additional articles missed by our search strategy, we cross-checked reference lists of all the studies included in this systematic review.

2.3.2. Selection of eligible studies

To select eligible studies (Figure 2.1), we followed the preferred reporting items for systematic reviews and meta-analyses (PRISMA) guidelines (Moher et al., 2009). Based on titles and abstracts, all records identified through the keyword search were assessed by two independent authors. Full text articles of studies selected as potentially relevant were thoroughly and independently assessed by two authors for final inclusion according to the listed criteria in Table 2.2.

2.3.3. Data extraction from each eligible study

Two authors independently performed data extraction from each eligible article using a standardized data extraction form. Any inconsistencies or disagreements between the two main reviewers were resolved by consensus, with a third reviewer requested to arbitrate on unresolved disagreements. We extracted the following information from each eligible study: country where the study was conducted, study type, age at which specimens were collected, duration of follow-up, methods used to define wheeze outcome, number of participants positive for wheeze, techniques used to detect faecal bacteria, faecal bacteria detected and odds ratio for the specific bacteria detected that were associated with wheezing illness (Table 2.3, 2.4).

2.3.4. Data analysis and synthesis

We performed narrative synthesis of included studies according to the research objectives and presented data in tables or figures. This included reporting on the total number of eligible studies, the main characteristics of the studies, data on the association between the faecal bacteria and the development of wheezing and recurrent wheezing, the factors affecting wheezing and the bacterial composition of faeces. We reported on the prevalence of specific bacteria, where appropriate.

Table 2.1. Search strategy performed in four electronic databases.

Database	Search mode	Keywords
Medline via PubMed	All fields	(microbiomic OR viromic OR microbiomics OR viromics OR microflora OR microbiome OR microbiota OR microbiomes OR microbiotas OR “human microbiotas” OR “human microbiomes” OR “gastrointestinal microbiome” OR “gut microbiotas” OR “gut microbiomes” OR “intestinal flora” OR “intestinal microflora” OR “digestive flora” OR “gut flora” OR metagenome OR metagenomic OR virome OR mycobiome OR mycobiomes OR mycobiota OR mycobiotas OR "fungal microbiome" OR “fungal microbiomes” OR “fungal microbiota” OR “fungal microbiotas”) AND (feces OR stool OR faeces OR faecal OR fecal) AND (“respiratory sounds” OR “respiratory sound” OR “lung sounds” OR “lung sound” OR wheez* OR asthma* OR “bronchial asthma” OR “asthma predictive index”)
Academic Search Premier, Africa-Wide Information and CINAHL via EBSCOHost	Boolean/Phrase	
Web of Science via Web of Knowledge	Topic	
Scopus via SciVerse	Article title, abstract, keywords	

Table 2.2. Eligibility criteria for the systematic review.

Inclusion criteria	Exclusion criteria
• Studies published through 19 August 2017 • Microbiological studies using faecal specimens • Cohort, cross-sectional or case-control studies • Studies carried out in infants	• Microbiological assessment of specimens other than faeces • Studies not reporting on wheeze • Assessment of the effect of antibiotic, pre- or probiotic treatment • Review articles, editorials, commentaries, or experimental studies

2.4. Results

2.4.1. Study selection

We identified 3688 articles from four electronic databases (Figure 2.1). Following the review of the titles and abstracts of the 3568 unduplicated articles from electronic databases, 3548 articles were excluded. Of the remaining 20 articles for which full-text readings were performed, only 9 articles fulfilled our inclusion criteria (Figure 2.1).

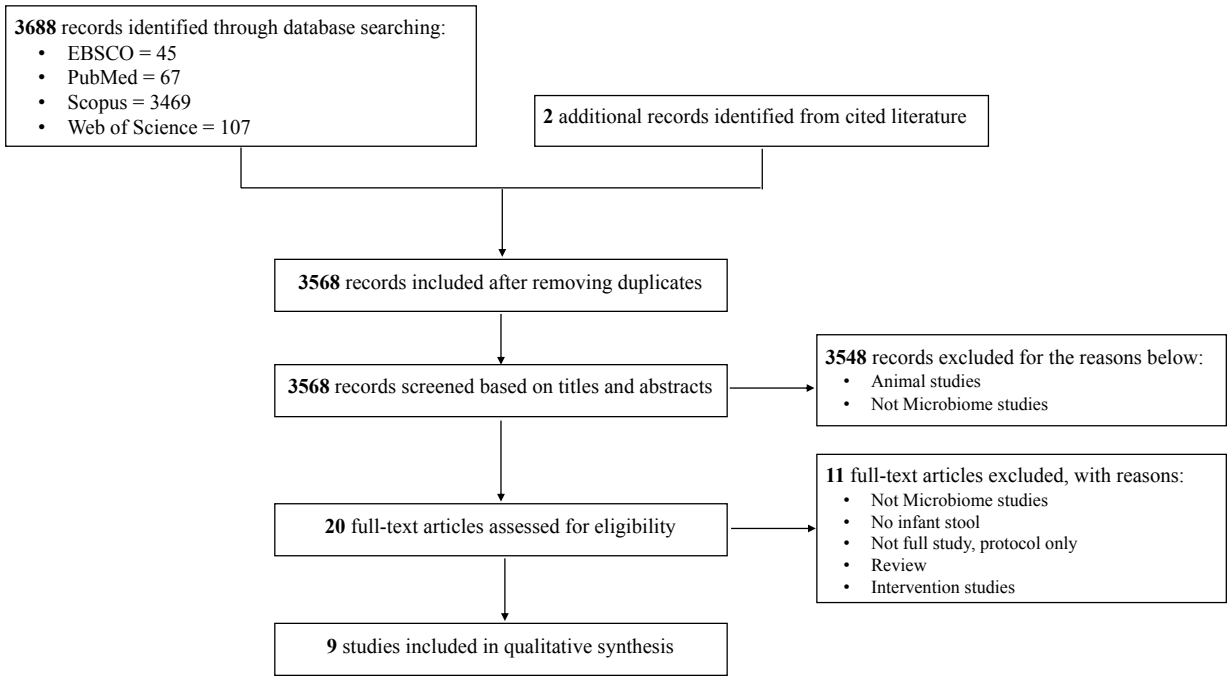


Figure 2.1. Flow diagram of systematic review process using the Preferred Reporting Items for Systematic Review and Meta-Analyses (PRISMA) guidelines.

2.4.2. Eligible microbiota studies that reported on single wheeze episode

2.4.2.1. Characteristics of studies

Table 2.3 summarizes the main characteristics of the 2 studies included in this systematic review that reported on the development of single wheezing episode only, and did not document recurrent wheezing. One of the studies was a prospective birth cohort whilst the other was just a prospective cohort study. The Japanese study reported on faecal microbial profiles of the infants

at 4 months whilst the Ethiopian study was at 3 years of age. The Japanese study was carried out in 2004 whilst the Ethiopian study was in 2011. Between the two studies, the wheezing definition was either based on doctors' assessment or parental reports. The Ethiopian and Japanese studies recorded the wheezing diagnosis at ages of three years and one year respectively (Figure 2.3). Faecal bacteria profiles of the two studies were characterized by culture-dependent techniques. None of the included studies reported on the use of either prebiotics or probiotics. None of the two studies on wheezing only demonstrated causal evidence for wheezing development. Amberbir and colleagues observed that the use of paracetamol and antibiotics in the child were significantly associated with wheezing (Amberbir *et al.*, 2011).

2.4.2.2. Association between the faecal bacteria and wheezing illness

The main focus of this review was the association between faecal bacteria and wheezing illness. The two studies that reported on single wheeze episodes did not find any significant association between the wheezing outcome and the faecal bacteria. Amberbir and colleagues investigated the association between wheezing and *Enterococci*, *Lactobacilli* and *Bifidobacteria*, but none were significantly associated (Amberbir *et al.*, 2011). Nambu and colleagues did not report on the odds ratio (OR), and did not observe any significant associations between the faecal bacteria and wheezing illness in infants (Nambu, Shintaku and Ohta, 2004).

2.4.3. Eligible microbiota studies that reported on recurrent wheezing

2.4.3.1. Characteristics of studies and risk factors affecting the development of wheezing

Table 2.4 summarizes the main characteristics of the 7 studies included in this systematic review. All of these 7 studies reported on recurrent wheezing, a subset of these studies also documented single wheezing episodes as well (Verhulst *et al.*, 2008; Van Nimwegen *et al.*, 2011; Gosalbes *et al.*, 2013; Arrieta *et al.*, 2015). All the studies reported were prospective birth cohort studies

and reported on faecal microbial profiles of the infants at different ages ranging from birth to 3 years. One of the studies included in this review looked at the faecal microbial profile at two separate time points in the same participants. Most of the eligible studies were conducted in Europe (86%, 6/7), between 2005 and 2013. Wheezing illness definition was heterogeneous and its diagnosis was based on doctors' assessment or parental reports. Most of the studies (67%) included in this review were collected in the first three months of life (Figure 2.2). Faecal bacterial screening in children less than five months was performed in 86% (6/7) of the studies. Faecal bacteria profiles were characterized by culture-dependent techniques in 14% (1/7) of the eligible studies. Culture-independent techniques used included quantitative real-time polymerase chain reaction (qPCR) (43%; 3/7), denaturing gradient gel electrophoresis (DGGE) in combination with PCR (14%; 1/7) (Murray, Tannock, Simon, Harmsen, Welling, Custovic, *et al.*, 2005), fluorescent in-situ hybridization (FISH) (14% ; 1/7) and massively parallel generation sequencing of 16S rRNA genes (29%; 2/7). None of the included studies reported on the use of either prebiotics or probiotics. Of the seven studies on recurrent wheezing only one (14%) demonstrated causal evidence for wheezing development. The study by Verhulst and colleagues observed a trend indicating the protective effect of breastfeeding on wheezing development (OR=0.54: 95% CI: 0.28 to 1.07) p=0.07, while the use of antibiotics and having male gender was associated with increased risk of wheezing. The study by Murray and colleagues recorded wheezing at the age of three years (Murray, 2005).

2.4.3.2. Association between the faecal bacteria and wheezing illness

Of the recurrent wheezing studies included in this review, 71% (5/7) of them reported the bacterial species that were significantly associated with the development of wheezing (Table 2.4). The remaining 29% of studies did not observe any significant associations between the bacteria and the development of wheezing. Arrieta and colleagues observed that four bacterial

genera (FLVR) were significantly reduced in the gut bacteria of wheezing infants, though they did not report the OR (Arrieta *et al.*, 2015). To prove causation, they investigated the impact of these four genera in a murine model of airway inflammation using a human faecal microbiota; some were supplemented with FLVR bacteria and others were not. Mice born to the parents supplemented with FLVR perpetuated the strains and exhibited reduced lung inflammation when compared to those without (Arrieta *et al.*, 2015). Colonization with *Bifidobacteria* and anaerobic bacteria at three weeks of age was associated with an increased risk of wheezing during the first year of life (OR = 3.65; 95% CI 1.21 to 11.0) and (OR = 1.14; 95% CI 1.02 to 1.28) respectively (Verhulst *et al.*, 2008). Some of the studies included did not observe any association between *Bifidobacterium* and wheezing (Murray, Tannock, Simon, Harmsen, Welling and Custovic, 2005a; Penders, Thijs, P. a van den Brandt, *et al.*, 2007; Amberbir *et al.*, 2011; Van Nimwegen *et al.*, 2011; Arrieta *et al.*, 2015). Verhulst and colleagues also observed that colonization with *Staphylococcus* was significantly associated with a decreased risk for wheezing (OR= 0.95; 95% CI 0.90 to 1.00) p=0.048 (Verhulst *et al.*, 2008). Conflicting data was also found in the studies included in this review. A study by a groups of researchers reported the taxon *Clostridium* as protective against wheezing (Verhulst *et al.*, 2008). However, three other studies suggested that the taxon *Clostridium* had more of a causative role in relation to wheezing in infancy (Penders, Thijs, P. A. van den Brandt, *et al.*, 2007; Penders *et al.*, 2008; Van Nimwegen *et al.*, 2011). The study by Verhulst and colleagues reported that colonization with the genus *Clostridia* at three weeks of life was significantly associated with a decreased risk for wheezing during the first year of life (OR= 0.53; 95% CI 0.30 to 0.92) (Verhulst *et al.*, 2008). However, three of the included studies reported that infants colonized with *Clostridium difficile* had a significantly increased risk of wheezing (43%, 3/7) (Penders, Thijs, P. a van den Brandt, *et al.*, 2007; Penders *et al.*, 2008; Van Nimwegen *et al.*, 2011). The Canadian study reported significant reductions in the presence of four bacterial genera (*Faecalibacterium*, *Lachnospira*, *Vellionella* and *Rothia*)

termed FLVR, in the faecal matter of the atopic wheezing children, but these same differences were not observed in wheezing only children (Arrieta *et al.*, 2015). The study by Arrieta and colleagues also confirmed that early microbial profile (3 months) is more suited to predicting wheezing later on in life, than the microbial profile at one year (Arrieta *et al.*, 2015).

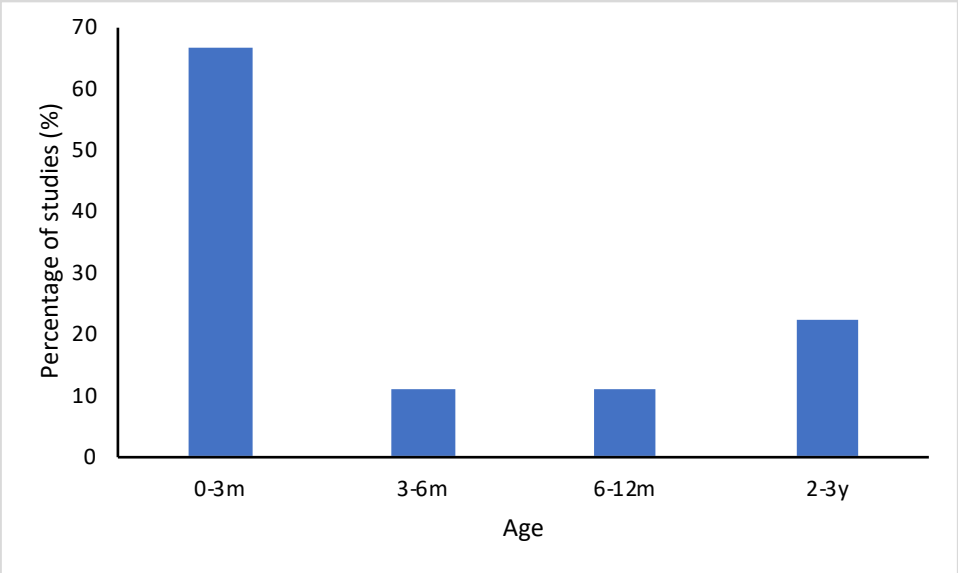


Figure 2.2. Age at which faecal samples were collected in all the nine studies under investigation.

Table 2.3. Characteristics and main findings of studies analyzing the association between faecal bacteria and the development of wheezing.

Country	Study design	Age at which specimens were collected	Age at which participants were followed up for wheeze	Wheeze diagnosis	Participants with wheeze % (n/N)	Techniques used to detect bacteria	Faecal bacteria detected	Risk associated with wheeze development			Citation
								OR (95% CI)	Cases	Controls	
Ethiopia	Prospective birth cohort	3 years	2-3 years	ISAAC questionnaires completed by parents	9.1 (80/876)	Culture-based	*Enterococci*	1.27 (0.72 - 2.24)	NR	NR	Amberbir et al., 2011
							Lactobacilli	1.19 (0.66 - 2.16)	NR	NR	
							Bifidobacteria	1.24 (0.63 - 2.47)	NR	NR	
						Antigen test	*Helicobacter pylori*	0.80 (0.46 - 1.38)	NR	NR	
Japan	Prospective cohort	4 months	1 year	Doctor diagnosed wheezing	16.7 (3/18)	Culture-based	*Lactobacilli*	NR	NR	NR	Nambu et al., 2004
							Bifidobacteria	NR	NR	NR	
							Clostridia	NR	NR	NR	

Table 2.4. Characteristics and main findings of studies analyzing the association between faecal bacteria and the development of recurrent wheezing.

Country	Study type	Age at which specimens were collected	Age at which participants were followed up for wheeze	Wheeze diagnosis	Participants positive for wheeze: % (n/N)	Techniques used to detect bacteria	Faecal bacteria detected	Risk associated with wheeze development			Citation
								OR (95% CI)	Cases	Controls	
Belgium	Prospective birth cohort	3 weeks	3 weeks	ISAAC questionnaires completed by parents	11.8	Culture-based	***Clostridium*¤**	**0.53 (0.30 - 0.92)**	NR	NR	Verhulst et al., 2008
			6 months		18.4		***Bifidobacterium***	**1.04 (0.99 - 1.09)**	NR	NR	
			12 months		23.5(154)		***Staphylococcus*¤**	**0.95 (0.90 – 1.00)**	NR	NR	
Canada	Prospective birth cohort	3 months	3 months	Doctor-diagnosed during API assessment & ISAAC questionnaires completed by parents	42.6 (136/319)	16S rRNA sequencing and quantitative PCR (qPCR)	***Faecalibacterium*¤**	NR	NR	NR	Arrieta et al., 2015
		1 year	6 months				***Lachnospira*¤**	NR	NR	NR	
			12 months				***Veillonella*¤**	NR	NR	NR	
			3 & 5 years				***Rothia*¤**	NR	NR	NR	
							***Bifidobacterium*¤**	NR	NR	NR	
Nether-lands	Prospective birth cohort	1 month	2 years	ISAAC questionnaires completed by parents	10 (957)	Quantitative real time PCR	*Bifidobacteria*	1.32 (0.85 - 2.06)	NR	NR	Penders et al., 2007
							Escherichia coli	1.92 (0.80 - 4.59)	NR	NR	
							Clostridium difficile	**1.75 (1.09 - 2.80)**	NR	NR	
							B. fragilis group	1.20 (0.66 - 2.18)	NR	NR	
							Lactobacilli	1.22 (0.77 - 1.93)	NR	NR	
Nether-lands	Prospective birth cohort	1 month	2 years	ISAAC questionnaires completed by parents	10 (957)	Quantitative real time PCR	**Toxigenic *Clostridium difficile***	**3.92 (1.62 - 9.44)**	NR	NR	Penders et al., 2008
							Non-toxigenic *Clostridium difficile*	1.48 (0.88 - 2.47)	NR	NR	

Nether-lands	Prospective birth cohort	1 month	7 months 12 months 24 months 4-5 years 6-7 years	Parentally reported wheezing at 7 months, 12 months, 24 months, 4-5 years, and 6-7 years	16.9 (420/2479 [ϕ]) 8.0 (154/1923[ʊ])	Quantitative real time PCR	***Clostridium difficile*** *Bifidobacterium* spp. *Escherichia coli* *Bacteroides fragilis* *Lactobacillus* spp.	**1.38 (1.05 - 1.81)**[¥] 1.96 (0.71 - 5.39)[¥] 1.22 (0.81 - 1.84)[¥] 0.89 (0.66 - 1.21)[¥] 0.96 (0.73 - 1.26)[¥]	NR NR NR NR NR	NR NR NR NR NR	Van Nimwegen et al., 2011
Spain	Prospective birth cohort	Birth	1 year 4 years	Parental reports on wheezing and medication used for wheezing during the fourth years of age	31.6 (6/20)	454 GS-FLX Titanium pyro-sequencing	Meconium microbiota type A[º] and type B[Ω]	0.62 (0.09 - 4.40)	NR	NR	Gosalbes et al. 2013
United Kingdom	Prospective birth cohort	3 years	3 years	At least 3 parentally reported wheezy episodes	50 (33/66)	PCR-DGGE	Lactic acid bacteria *Bifidobacterium* spp.	NR NR	12 30	11 31	Murray et al., 2005
						FISH	*Bifidobacterium adolescentis* *Bifidobacterium bifidum* *Bifidobacterium longum* *Bifidobacterium catenulatum* *Bifidobacterium pseudocatenulatum*	NR NR NR NR NR	28 4 13 1 2	30 5 12 1 2	

Significant results are highlighted in boldface

API: Asthma Predictive Index; CI: Confidence interval; DGGE: Denaturing gradient gel electrophoresis; FISH: Fluorescent in-situ hybridization; ISAAC: International Study of Asthma and Allergies in children; n: Number tested positive; N: Total number tested; NR: Not reported; OR: Odds ratio; PCR: Polymerase chain reaction

[ɣ] p = 0.01, but no confidence interval reported

[§] *Bacteroides fragilis*, *Bacteroides finegoldii* and *Bacteroides thetaiotaomicron*

[†] *Ruminococcus productus, Ruminococcus hansenii* and *Clostridium* spp.

[‡] *Eubacterium contortum*, *Clostridium oroticum*, and *Ruminococcus torques*

[º] Family *Enterobacteriaceae* as the most abundant bacterial taxon

[Ω] Families *Leuconostocaceae, Enterococcaceae* and *Streptococcaceae* as the most abundant bacterial taxa

[¥] Generalized estimating equations of wheeze measured at 7, 12 and 24 months and 4-5 and 6-7 years

[ϕ] Prevalence of wheezing at 12 months

[ʊ] Prevalence of wheezing at 6-7 years

[Ω] No significant characteristics of intestinal microflora in the wheezing case

[¤] Bacteria associated with reduced risk of wheezing in infants

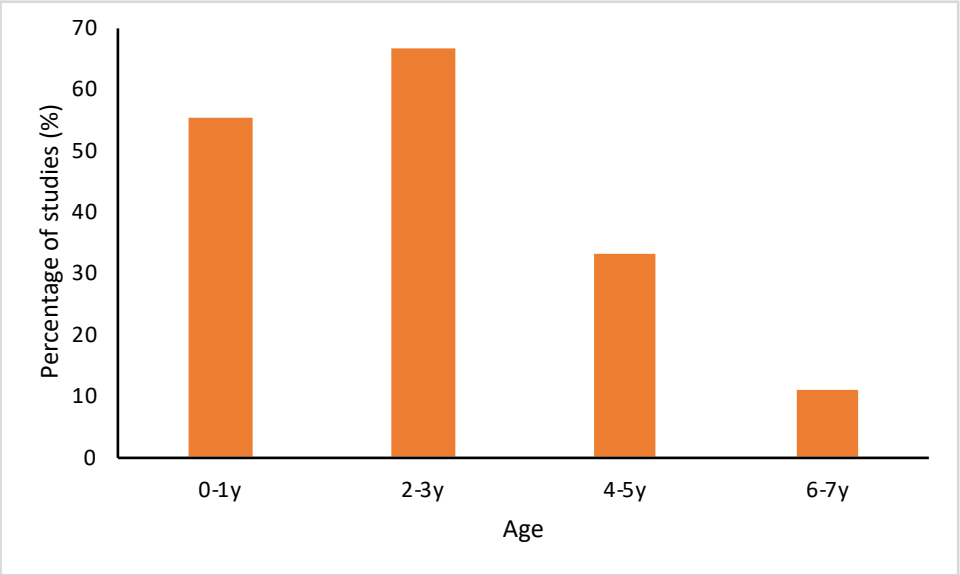


Figure 2.3. Age at which wheezing was reported in all the nine studies under investigation.
There is a total of nine studies included which reported wheezing at one or more different time periods. The time periods were 0-1 years, 2-3 years, 4-5 years and 6-7 years.

2.5. Discussion

The main finding of this systematic review was the association observed between the four FLVR bacterial genera and the development of atopic wheezing in infants at high risk of developing asthma (Arrieta *et al.*, 2015). All four bacterial genera were found to be significantly reduced at three months of age in the faeces of atopic wheezing children in the Canadian study, suggesting that they play a protective role in preventing wheezing illness (Arrieta *et al.*, 2015). The murine lung inflammation research by the Canadian group showed that supplementation with the FLVR bacteria significantly reduced the inflammation in the mice (Arrieta *et al.*, 2015). This supplementation can be used as a potential preventative method against the development of asthma later on in life, in children at risk. The identification of these key bacterial genera opens up the idea of a more focused research approach on select bacterial types in the development of wheezing in children. If the same results are observed all over the world, these bacteria could be

used for therapeutic interventions to restore balance to the gut bacteria in infants likely to develop wheezing.

Three studies from the same cohort observed an association between *C. difficile* and the increased risk of wheezing illness in young children (Penders, Thijs, P. a van den Brandt, *et al.*, 2007; Penders *et al.*, 2008; Van Nimwegen *et al.*, 2011). This finding is supported by Woodcock and colleagues who showed that *C. difficile* IgG antibody levels were significantly higher in allergic infants with recurrent wheeze compared to non-wheezy infants (Woodcock *et al.*, 2002). It is also hypothesized that the role of faecal *C. difficile* in the occurrence of wheezing is to disrupt the infant's immune system development (Azad and Kozyrskyj, 2012). The use of antibiotics in infants prompts the emergence of *C. difficile* and the suppression of commensal bacteria which would be involved in the development of the immune system (Wilson, 1993; Chang *et al.*, 2008; Murgas Torrazza and Neu, 2011; Azad and Kozyrskyj, 2012; Theriot *et al.*, 2014). This unbalanced/dysbiotic gut then increases the probability of developing wheezing in the infant, since the protective bacteria would be absent. Though the specific mechanism that links the gut microbiota and infant wheezing is unclear, there are some recent studies that have identified a significant amount of crosstalk between gut microbiota and host immune cells (Vital *et al.,* 2015, Fosythe 2011). Dendritic cells present in the gut, and in other mucosal site like the lungs, transport bacterial antigens to lymph nodes where they can influence local immune responses (Hill & Artis 2011, Salzman 2011, Samuelson *et al., 2015*). This interaction therefore significantly increases the risk of development of respiratory diseases (Hill & Artis 2011). The paper by Claassen-Weitz and colleagues summarizes the proposed gut microbiota and infant wheeze mechanisms well (Claassen-Weitz *et al.,* 2016).

Previous studies have reported an association between prenatal and early antibiotic exposure and the development of wheezing (Lapin *et al.*, 2015; Popovic *et al.*, 2016; Han *et al.*, 2017). Alternatively, *C. difficile* may itself be a specific risk factor for the development of wheezing, independent of antibiotic use. Van Nimwegen and colleagues used mediation analysis to confirm the intermediary role by *C. difficile* in the association of mode and place of delivery and development of atopy (Van Nimwegen *et al.*, 2011). The reason was that gut colonization by *C. difficile* might act as an intervening factor between birth mode and development of asthma and atopic diseases (Van Nimwegen *et al.*, 2011). *C. difficile* mediation was found for vaginal hospital delivery versus caesarean section delivery, as well as vaginal home delivery in association with asthma (Van Nimwegen *et al.*, 2011). Infants delivered vaginally at home would potentially be colonized by *C. difficile* since they were likely to be exposed to minimal or no medication, when compared to those delivered in hospital (Van Nimwegen *et al.*, 2011). The above findings strengthen the association caused by mode of delivery on the development of atopic outcomes. Furthermore, another included study demonstrated the importance of genotyping in association studies by investigating the difference between toxigenic and non-toxigenic *C. difficile* in wheeze development (Penders *et al.*, 2008). Penders and colleagues demonstrated that recurrent wheezing was only associated with toxin-positive *C. difficile* and not its toxin-negative counterpart which is also found in same environment (Penders *et al.*, 2008). Therefore it is necessary for association studies to go a step further than describing bacteria to genus and species level, in order to determine the specific bacterial type responsible for development of wheezing.

Three studies included in this review investigated the association between wheezing illness and colonization by *Bifidobacteria* in infants (Verhulst *et al.*, 2008; Amberbir *et al.*, 2011; Arrieta *et al.*, 2015). Two of these studies concluded that colonization by *Bifidobacteria* was associated

with increased probability of wheezing (Verhulst *et al.*, 2008; Amberbir *et al.*, 2011). However, the study by Arrieta and colleagues did not observe any differences in the proportions of *Bifidobacteria* between controls and atopic wheezing children (Arrieta *et al.*, 2015). A possible explanation for the differences could be attributed to breastfeeding, since *Bifidobacteria* colonization has been shown to be associated with breastfeeding (Hesla *et al.*, 2014; Parra-Llorca *et al.*, 2018). Breastfeeding has also been shown to be protective against wheezing by multiple mechanisms some of which include modulating microflora and promoting gastrointestinal mucosa maturation (Oddy, 2004). Therefore, more research needs to be carried out to fully explain the effect of breastfeeding with regards to wheezing in different communities. The use of probiotics, prebiotics, or symbiotics for the modulation of gastrointestinal microbiome was not reported in studies included in this review. Previous findings from three independent meta-analysis studies suggest that probiotics do not significantly reduce wheezing in infants (Azad *et al.*, 2013; Elazab *et al.*, 2013; Zuccotti *et al.*, 2015). Similarly, a meta-analysis based on two studies suggested the evidence for the protective role of prebiotics in the development of wheezing illness was low (Osborn and Sinn, 2013). More strain specific research needs to be carried out in this field to determine the influence of prebiotics and probiotics in the development of wheezing, since some probiotic benefits have been observed to be strain related (Ling *et al.*, 2015).

Most of the studies included in this review assessed the GIT microbial community by use of culture-dependent or culture-independent methods which are limited in their strength of providing a comprehensive picture of the contribution of specific microbial community groups in the occurrence of wheezing illness. Only two studies (conducted in Spain and Canada) used massively parallel sequencing to assess this issue (Gosalbes *et al.*, 2013; Arrieta *et al.*, 2015). Using NGS, the Canadian birth cohort study was able to successfully determine differentially

abundant genera between infants that wheezed and those that did not. They singled out the FLVR bacterial genera, that were significantly reduced in the wheezing children at the age of three months (Arrieta *et al.*, 2015). However, at one year of age only two of the bacterial genera (*Veillonella* and *Lachnospira*) had significant differences between the controls and the atopic wheezing children (Arrieta *et al.*, 2015). Though 16S sequencing is very informative, it can only provide information with confidence up to the genus level, other methods would be needed to obtain species-level information. Since the early 2000's the cost of sequencing has been decreasing dramatically, thus future studies can employ strain-specific bacterial typing by using whole-genome shotgun-based method which offers higher and more reliable resolution of the composition of the microbiota at species-level (Morgan and Huttenhower, 2014; Van Dijk et al., 2014). Furthermore, more studies have been investigating the metabolic role played by bacteria in pathogenesis of diseases, since the bacteria also play a role in the metabolic processes of the human body (Morgan and Huttenhower, 2014; Arrieta *et al.*, 2015; Cardet, Johns and Savage, 2015). The Canadian study found that faecal acetate was significantly reduced in atopic wheezing infants in comparison to the controls (Arrieta *et al.*, 2015). This is significant as acetate, a short-chain fatty acid (SCFA), is a microbe-derived metabolite. Acetate, propionate and butyrate are a few examples of SCFA that have beneficial anti-inflammatory properties in human subjects in relation to various health conditions such as Crohn's diseases and atopic wheezing (Morgan and Huttenhower, 2014; Arrieta *et al.*, 2015). Cardet and colleagues observed that bacteria derived metabolites were involved in the health state in relation to asthma and development of non-asthmatic wheeze (Cardet, Johns and Savage, 2015). Cardet and colleagues also observed that subjects presenting with higher lignan metabolite enterolactone had reduced diagnosis of current asthma and non-asthmatic wheeze (Cardet, Johns and Savage, 2015). In addition, other studies have also supported the anti-inflammatory role played by the other metabolites including

prostaglandin F2 alpha and allantoin in respiratory diseases (Skoner *et al.*, 1989; Chiu *et al.*, 2018).

One of the studies only observed differences in a few taxa when a comparison of the relative abundance was carried out at three months, versus one year (Arrieta *et al.*, 2015). They hypothesize that these differences were not detected at one year because the 'critical window' of the first one hundred days of life, in which gut microbial dysbiosis associated with wheezing illness can be detected had passed (Van Der Velden *et al.*, 2001; Arrieta *et al.*, 2015) . After the hypothesized 'critical window' has passed, no major differences can be observed between the bacterial profiles of the wheezers and non-wheezers (Van Der Velden *et al.*, 2001; Arrieta *et al.*, 2015). More studies are required to confirm the 'critical window' when it comes to wheeze development in infants from different regions. Furthermore, additional independent longitudinal studies with multiple sampling points that fall within the critical window are required to better define the progression, with regards to the relative taxa abundance of the bacteria associated with wheezing, including the FLVR bacteria. These studies would provide more insightful missing information on this topic.

Limitations to this systematic review are that most of the eligible studies did not conduct rigorous statistical analysis to account for confounding factors that may be associated independently with wheeze, as well as influencing faecal bacteria profiles. The majority of the studies were conducted in developed countries except for the one study in Ethiopia. Therefore, it would be of great interest to perform similar studies in many low- and middle-income countries. This is of particular interest, since differences in microbial communities are reported between developed and developing countries (De Filippo et al., 2010; Grześkowiak et al., 2012; Yatsunenko et al., 2012; Lin et al., 2013; Ou et al., 2013; Suzuki and Worobey, 2014). The contribution of the

bacterial community in the occurrence of wheezing in Africa is currently unknown; therefore, more studies are warranted in different populations across the continent. Furthermore, the definitions of outcomes assessed in this review were heterogeneous and observed at different ages (Murray et al., 2005; Penders et al., 2007, 2008; Verhulst et al., 2008; Amberbir et al., 2011; Van Nimwegen et al., 2011; Gosalbes et al., 2013). For example, Arrieta and colleagues defined their main outcome as atopic wheezing, which was the presence of a positive allergen skin prick test and one or more wheezing episodes in the first year of life (Arrieta *et al.*, 2015). On the other hand, Van Nimwegen and colleagues defined their outcome as wheezing according to the International Study of Asthma and Allergies in Childhood (ISAAC) questionnaire (Van Nimwegen *et al.*, 2011). The different outcome definitions do not allow for a full comparative analysis since the outcomes exhibit varying degrees of wheezing illness intensity. According to previous literature, infant wheezing is more prevalent in the first three years of life (Martinez *et al.*, 1995). Most of the studies included in this review fall in this range (Nambu, Shintaku and Ohta, 2004; Penders, Thijs, P. a van den Brandt, *et al.*, 2007; Penders *et al.*, 2008; Verhulst *et al.*, 2008; Van Nimwegen *et al.*, 2011; Gosalbes *et al.*, 2013; Arrieta *et al.*, 2015). The studies with infants three years and older have a potential to produce less information, as the prime period in which wheezing occurs would have passed (Murray, Tannock, Simon, Harmsen, Welling and Custovic, 2005a; Amberbir *et al.*, 2011). In addition, studies that have postulated about the 'critical window', mention it being in the first 100 days of life, this is a time period in which microbial dysbiosis can be linked to wheezing or allergic disease (Russell *et al.*, 2012; Arrieta *et al.*, 2015). Therefore, bacterial investigations outside this 100 day age range would most likely not be able to show the same information with regards to microbial dysbiosis that can be linked to wheezing illness.

2.6. Conclusion

Few studies have reported on the significant associations between faecal bacteria and wheezing. Most of these studies had small sample sizes and used target-based approaches for studying the bacterial community. Future studies would require large, longitudinal birth cohort studies that use high-resolution experimental methods for microbiota analysis. This exhaustive analysis would enable researchers to extensively understand the role of faecal microbial profiles in the development of wheezing illness in both developed and developing countries. Finally, further research from other cohorts assessing the association between the FLVR bacteria, *C. difficile* colonization and wheezing illness in young children is required. The reason would be all the publications reported here were generated from the same cohort and were thus not independent.

2.7. References

Amberbir, a. *et al.* (2011) 'Effects of Helicobacter pylori, geohelminth infection and selected commensal bacteria on the risk of allergic disease and sensitization in 3-year-old Ethiopian children', *Clinical and Experimental Allergy*, 41(10), pp. 1422–1430. doi: 10.1111/j.1365-2222.2011.03831.x.

Arden, K. E. *et al.* (2006) 'Frequent detection of human rhinoviruses, paramyxoviruses, coronaviruses, and bocavirus during acute respiratory tract infections.', *Journal of medical virology*, 78(9), pp. 1232–1240. doi: 10.1002/jmv.20689.

Arrieta, M. *et al.* (2015) 'Early infancy microbial and metabolic alterations affect risk of childhood asthma', *Science Translational Medicine*, 7(307), p. 307ra152-307ra152. doi: 10.1126/scitranslmed.aab2271.

Azad, M. B. *et al.* (2013) 'Probiotic supplementation during pregnancy or infancy for the prevention of asthma and wheeze: systematic review and meta-analysis.', *BMJ (Clinical research ed.)*, 347, p. f6471.

Azad, M. B. and Kozyrskyj, A. L. (2012) 'Perinatal programming of asthma: The role of gut microbiota', *Clinical and Developmental Immunology*, doi: 10.1155/2012/932072.

Been, J. V *et al.* (2014) 'Preterm birth and childhood wheezing disorders: a systematic review and meta-analysis.', *PLoS medicine*, 11(1), p. e1001596. doi: 10.1371/journal.pmed.1001596.

Bisgaard, H. *et al.* (2011) 'Reduced diversity of the intestinal microbiota during infancy is associated with increased risk of allergic disease at school age', *Journal of Allergy and Clinical Immunology*, 128. doi: 10.1016/j.jaci.2011.04.060.

Bisgaard, H. and Szefler, S. (2007) 'Prevalence of asthma-like symptoms in young children', *Pediatric Pulmonology*, 42(8), pp. 723–728. doi: 10.1002/ppul.20644.

Cardet, J. C., Johns, C. B. and Savage, J. H. (2015) 'Bacterial metabolites of diet-derived lignans and isoflavones inversely associate with asthma and wheezing', *Journal of Allergy and Clinical Immunology*, 135(1), pp. 267–269. doi: 10.1016/j.jaci.2014.07.035.

Chang, J. Y. *et al.* (2008) 'Decreased diversity of the fecal microbiome in recurrent Clostridium difficile-associated diarrhea.', *The Journal of infectious diseases*, 197(3), pp. 435–8. doi: 10.1086/525047.

Chiu, C.-Y. *et al.* (2018) 'Longitudinal urinary metabolomic profiling reveals metabolites for asthma development in early childhood', *Pediatric Allergy and Immunology*, (April), pp. 496–503. doi: 10.1111/pai.12909.

Claassen-Weitz et al. (2016) ' Current Knowledge and Future Research Directions on Fecal Bacterial Patterns and Their Association with Asthma', Frontiers in Microbiology, 29(7), pp838. doi: 10.3389/fmicb.2016.00838. eCollection 2016.

Van Dijk, E. L. *et al.* (2014) 'Ten years of next-generation sequencing technology', *Trends in Genetics*, 30(9), pp. 418–426. doi: 10.1016/j.tig.2014.07.001.

Elazab, N. *et al.* (2013) 'Probiotic administration in early life, atopy, and asthma: A meta-analysis of clinical trials', *Pediatrics*, 132(3), pp. e666–e676. doi: 10.1542/peds.2013-0246.

Esposito, S. *et al.* (2014) 'Impact of air pollution on respiratory diseases in children with recurrent wheezing or asthma.', *BMC pulmonary medicine*, 14, p. 130. doi: 10.1186/1471-2466-14-130.

Ferguson, A. C., Whitelaw, M. and Brown, H. (1992) 'Correlation of bronchial eosinophil and mast cell activation with bronchial hyperresponsiveness in children with asthma.', *The Journal of allergy and clinical immunology*, 90(4 Pt 1), pp. 609–613.

De Filippo, C. *et al.* (2010) 'Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa.', *Proceedings of the National Academy of Sciences of the United States of America*, 107(33), pp. 14691–6. doi: 10.1073/pnas.1005963107.

Forsythe, P. (2011). Probiotics and lung diseases. *Chest* 139, 901–908. doi: 10.1378/chest.10-1861

Friedman, N. J. and Zeiger, R. S. (2005) 'The role of breast-feeding in the development of allergies and asthma.', *The Journal of allergy and clinical immunology*, 115(6), pp. 1238–1248. doi: 10.1016/j.jaci.2005.01.069.

Gosalbes, M. J. *et al.* (2013) 'Meconium microbiota types dominated by lactic acid or enteric bacteria are differentially associated with maternal eczema and respiratory problems in infants', *Clinical & Experimental Allergy*, 43(2), pp. 198–211. doi: 10.1111/cea.12063.

Grześkowiak, Ł. *et al.* (2012) 'Distinct Gut Microbiota in Southeastern African and Northern European Infants', *Journal of Pediatric Gastroenterology and Nutrition*, 54(6), pp. 812–816. doi: 10.1097/MPG.0b013e318249039c.

Han, Y.-Y. *et al.* (2017) 'Antibiotic use in early life, rural residence, and allergic diseases in Argentinean children.', *The journal of allergy and clinical immunology*, 5(4), p. 1112–1118.e2. doi: 10.1016/j.jaip.2016.12.025.

Henderson, J. *et al.* (2008) 'Associations of wheezing phenotypes in the first 6 years of life with atopy, lung function and airway responsiveness in mid-childhood', *Thorax*, 63(11), pp. 974–980. doi: 10.1136/thx.2007.093187.

Hesla, H. M. *et al.* (2014) 'Impact of lifestyle on the gut microbiota of healthy infants and their mothers-the ALADDIN birth cohort.', *FEMS microbiology ecology*, 90(3), pp. 791–801. doi: 10.1111/1574-6941.12434.

Hill, D. A., and Artis, D. (2010). Intestinal bacteria and the regulation of immune cell homeostasis. *Annual Review of Immunol*ogy 28, 623–667. doi: 10.1146/annurevimmunol-030409-101330

Ip, S. *et al.* (2007) 'Breastfeeding and maternal and infant health outcomes in developed countries.', *Evidence report/technology assessment*, (153), pp. 1–186.

Jackson, D. J. *et al.* (2012) 'Evidence for a causal relationship between allergic sensitization and rhinovirus wheezing in early life', *American Journal of Respiratory and Critical Care Medicine*, 185(3), pp. 281–285. doi: 10.1164/rccm.201104-0660OC.

Johnston, S. L. *et al.* (1996) 'The relationship between upper respiratory infections and hospital admissions for asthma: a time-trend analysis.', *American journal of respiratory and critical care medicine*, 154(3 Pt 1), pp. 654–660. doi: 10.1164/ajrccm.154.3.8810601.

Kotaniemi-Syrjanen, A. *et al.* (2003) 'Rhinovirus-induced wheezing in infancy--the first sign of childhood asthma?', *The Journal of allergy and clinical immunology*, 111(1), pp. 66–71.

Kusel, M. M. H. *et al.* (2007) 'Early-life respiratory viral infections, atopic sensitization, and risk of subsequent development of persistent asthma', *Journal of Allergy and Clinical Immunology*, 119, pp. 1105–1110. doi: 10.1016/j.jaci.2006.12.669.

Kwong, C. G. and Bacharier, L. B. (2017) 'Microbes and the role of antibiotic treatment for

wheezy lower respiratory tract illnesses in preschool children.', *Current allergy and asthma reports*, 17(5), p. 34. doi: 10.1007/s11882-017-0701-6.

Lapin, B. *et al.* (2015) 'Relationship between prenatal antibiotic use and asthma in at-risk children.', *Annals of allergy, asthma & immunology : official publication of the American College of Allergy, Asthma, & Immunology*, 114(3), pp. 203–207. doi: 10.1016/j.anai.2014.11.014.

Lau, S. *et al.* (2003) 'Transient early wheeze is not associated with impaired lung function in 7-yr-old children', *European Respiratory Journal*, 21(5), pp. 834–841. doi: 10.1183/09031936.03.00037203.

Lin, A. *et al.* (2013) 'Distinct distal gut microbiome diversity and composition in healthy children from Bangladesh and the United States', *PLoS ONE*, 8(1). doi: 10.1371/journal.pone.0053838.

Ling, Z. *et al.* (2015) 'Clostridium butyricum combined with Bifidobacterium infantis probiotic mixture restores fecal microbiota and attenuates systemic inflammation in mice with antibiotic-associated diarrhea.', *BioMed research international*, 2015, p. 582048. doi: 10.1155/2015/582048.

Liu, L. *et al.* (2017) 'Association between rhinovirus wheezing illness and the development of childhood asthma: a meta-analysis.', *BMJ open*, 7(4), p. e013034. doi: 10.1136/bmjopen-2016-013034.

Lodge, C. J. *et al.* (2015) 'Breastfeeding and asthma and allergies: a systematic review and meta-analysis.', *Acta paediatrica (Oslo, Norway : 1992)*, 104(467), pp. 38–53. doi: 10.1111/apa.13132.

Ly, N. P. *et al.* (2006) 'Recurrent wheeze in early childhood and asthma among children at risk for atopy.', *Pediatrics*, 117(6), pp. e1132–e1138. doi: 10.1542/peds.

Martinez, F. D. *et al.* (1995) 'Asthma and wheezing in the first six years of life', 332(3).

Mebrahtu, T. F. *et al.* (2015) 'Birth weight and childhood wheezing disorders: a systematic review and meta-analysis.', *Journal of epidemiology and community health*, 69(5), pp. 500–508. doi: 10.1136/jech-2014-204783.

Moher, D. *et al.* (2009) 'Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement', *PLoS medicine*, 6(7), p. e1000097.

Morgan, X. C. and Huttenhower, C. (2014) 'Meta'omic analytic techniques for studying the intestinal microbiome', *Gastroenterology*, 146(6), p. 1437–1448.e1. doi: 10.1053/j.gastro.2014.01.049.

Murgas Torrazza, R. and Neu, J. (2011) 'The developing intestinal microbiome and its relationship to health and disease in the neonate.', *Journal of perinatology : official journal of the California Perinatal Association*, 31 Suppl 1, pp. S29-34. doi: 10.1038/jp.2010.172.

Murray, C. S., Tannock, G. W., Simon, M. A., Harmsen, H. J. M., Welling, G. W., Custovic, A., *et al.* (2005) 'Fecal microbiota in sensitized wheezy and non-sensitized non-wheezy children: a nested case–control study', *Clinical & Experimental Allergy*, 35(6), pp. 741–745. doi: 10.1111/j.1365-2222.2005.02259.x.

von Mutius, E. (2001) 'Infection: friend or foe in the development of atopy and asthma? The epidemiological evidence', *The European respiratory journal*, 18(5), pp. 872–81.

Nambu, M., Shintaku, N. and Ohta, S. (2004) 'Intestinal microflora at 4 months of age and the development of allergy', *Allergology International*, 53(2), pp. 121–126. doi: 10.1111/j.1440-

1592.2004.00315.x.

Neuman, A. *et al.* (2014) 'Infant wheeze, comorbidities and school age asthma.', *Pediatric allergy and immunology : official publication of the European Society of Pediatric Allergy and Immunology*, 25(4), pp. 380–386. doi: 10.1111/pai.12223.

Van Nimwegen, F. a. *et al.* (2011) 'Mode and place of delivery, gastrointestinal microbiota, and their influence on asthma and atopy', *Journal of Allergy and Clinical Immunology*, 128(5), pp. 948–955. doi: 10.1016/j.jaci.2011.07.027.

Oddy, W. H. (2004) 'A review of the effects of breastfeeding on respiratory infections, atopy, and childhood asthma.', *The Journal of asthma : official journal of the Association for the Care of Asthma*, 41(6), pp. 605–621.

Oddy, W. H. *et al.* (2004) 'The relation of breastfeeding and body mass index to asthma and atopy in children: A prospective cohort study to age 6 years', *American Journal of Public Health*, 94(9), pp. 1531–1537. doi: 10.2105/AJPH.94.9.1531.

Osborn, D. and Sinn, J. (2013) 'Prebiotics in infants for prevention of allergy (Review)', *The Cochrane database for Systematic Reviews*, (3).

Ou, J. . *et al.* (2013) 'Diet, microbiota, and microbial metabolites in colon cancer risk in rural Africans and African Americans', *American Journal of Clinical Nutrition*, 98(1), pp. 111–120. doi: 10.3945/ajcn.112.056689.

Parra-Llorca, A. *et al.* (2018) 'Preterm gut microbiome depending on feeding type: significance of donor human milk.', *Frontiers in microbiology*, 9, p. 1376. doi: 10.3389/fmicb.2018.01376.

Penders, J., Thijs, C., van den Brandt, P. a, *et al.* (2007) 'Gut microbiota composition and development of atopic manifestations in infancy: the KOALA Birth Cohort Study.', *Gut*, 56(5), pp. 661–667. doi: 10.1136/gut.2006.100164.

Penders, J. *et al.* (2008) 'Toxigenic and non-toxigenic Clostridium difficile : determinants of intestinal colonisation and role in childhood atopic manifestations', *Gut*, 57(7), pp. 1025–1026. doi: 10.1136/gut.2007.143214.

Popovic, M. *et al.* (2016) 'Prenatal exposure to antibiotics and wheezing in infancy: a birth cohort study.', *The European respiratory journal*, 47(3), pp. 810–817. doi: 10.1183/13993003.00315-2015.

Rosas-Salazar, C. *et al.* (2018) 'Nasopharyngeal Lactobacillus is associated with a reduced risk of childhood wheezing illnesses following acute respiratory syncytial virus infection in infancy.', *The Journal of allergy and clinical immunology*. doi: 10.1016/j.jaci.2017.10.049.

Russell, S. L. *et al.* (2012) 'Early life antibiotic-driven changes in microbiota enhance susceptibility to allergic asthma.', *EMBO reports*, 13(5), pp. 440–7. doi: 10.1038/embor.2012.32.

Salzman, N. H. (2011). Microbiota-immune system interaction: an uneasy alliance. *Current Opinion in Microbiology* 14, 99–105. doi: 10.1016/j.mib.2010.09.018

Samuelson, D. R., Welsch, D. A., and Shellito, J. E. (2015). Regulation of lung immunity and host defense by the intestinal microbiota. *Frontiers in Microbiol*ogy 6:1085. doi: 10.3389/fmicb.2015.01085

De Schutter, I. *et al.* (2012) 'In young children, persistent wheezing is associated with bronchial bacterial infection: A retrospective analysis', *BMC Pediatrics*. ???, 12(1), p. 1. doi: 10.1186/1471-2431-12-83.

Sigurs, N. *et al.* (2010) 'Asthma and allergy patterns over 18 years after severe RSV bronchiolitis in the first year of life.', *Thorax*, 65(12), pp. 1045–1052. doi: 10.1136/thx.2009.121582.

Skoner, D. P. *et al.* (1989) 'Increases in plasma concentrations of a prostaglandin metabolite in acute airway obstruction.', *Archives of disease in childhood*, 64(8), pp. 1112–1117.

Stevens, C. A. *et al.* (2003) 'The economic impact of preschool asthma and wheeze.', *The European respiratory journal*, 21(6), pp. 1000–1006.

Suzuki, T. A. and Worobey, M. (2014) 'Geographical variation of human gut microbial composition Geographical variation of human gut microbial composition', *Biology letters*, 10(2).

Ta, L. D. H. *et al.* (2018) 'Establishment of the nasal microbiota in the first 18 months of life: Correlation with early-onset rhinitis and wheezing.', *The Journal of allergy and clinical immunology*, 142(1), pp. 86–95. doi: 10.1016/j.jaci.2018.01.032.

Taussig, L. M. *et al.* (2003) 'Tucson children's respiratory study: 1980 to present', *Journal of Allergy and Clinical Immunology*, 111(4), pp. 661–675. doi: 10.1067/mai.2003.162.

Theriot, C. *et al.* (2014) 'Antibiotic-induced shifts in the mouse gut microbiome and metabolome increase susceptibility to Clostridium difficile infection', *Nature Communications*, 5(3114).

Ungar, W. J. and Coyte, P. C. (2001) 'Prospective study of the patient-level cost of asthma care in children.', *Pediatric pulmonology*, 32(2), pp. 101–108.

Vanker, A., Gie, R. P. and Zar, H. J. (2018) 'Early-life exposures to environmental tobacco smoke and indoor air pollution in the Drakenstein Child Health Study: Impact on child health', *South African Medical Journal*, 108(2), p. 71. doi: 10.7196/SAMJ.2018.v108i2.13088.

Van Der Velden, V. H. J. *et al.* (2001) 'Selective development of a strong Th2 cytokine profile in high-risk children who develop atopy: Risk factors and regulatory role of IFN-γ, IL-4 and IL-10', *Clinical and Experimental Allergy*, 31(7), pp. 997–1006. doi: 10.1046/j.1365-2222.2001.01176.x.

Verhulst, S. L. *et al.* (2008) 'A longitudinal analysis on the association between antibiotic use, intestinal microflora, and wheezing during the first year of life.', *The Journal of asthma : official journal of the Association for the Care of Asthma*, 45, pp. 828–832. doi: 10.1080/02770900802339734.

Vital, M., Harkema, J. R., Rizzo, M., Tiedje, J., and Brandenberger, C. (2015). Alterations of the murine gut microbiome with age and allergic airway disease. *Journal of Immunology Res*earch 2015, 1–8. doi: 10.1155/2015/892568

Wilson, K. H. (1993) 'The microecology of Clostridium difficile.', *Clinical infectious diseases*, 16(Suppl. 4), pp. S214-8.

Woodcock, A. *et al.* (2002) 'Clostridium difficile, atopy and wheeze during the first year of life', *Pediatr Allergy Immunol*, 13(3), pp. 357–360.

Yatsunenko, T. *et al.* (2012) 'Human gut microbiome viewed across age and geography', *Nature*, 486(7402), pp. 222–227. doi: 10.1038/nature11053.

Zhao, D. *et al.* (2015) 'Prenatal antibiotic use and risk of childhood wheeze/asthma: A meta-analysis.', *Pediatric allergy and immunology : official publication of the European Society of Pediatric Allergy and Immunology*, 26(8), pp. 756–764. doi: 10.1111/pai.12436.

Zuccotti, G. *et al.* (2015) 'Probiotics for prevention of atopic diseases in infants: Systematic review and meta-analysis', *Allergy: European Journal of Allergy and Clinical Immunology*, 70(11), pp. 1356–1371. doi: 10.1111/all.12700.

Chapter 3 - Association between Faecal Bacteria and the Development of Recurrent Wheezing among Infants within the Drakenstein Child Health Study

3.1. Abstract

Background: Multiple studies have investigated the association between infant stool microbiota and wheezing illness. However, these studies have not been conclusive due to limitations such as targeted experimental techniques and small sample sizes. Therefore, to gain greater insight into the association between wheezing illness and the stool microbiota, we conducted a prospective, longitudinal study that investigated recurrent wheezing in infants using next-generation sequencing techniques.

Methods: The infants were from the Drakenstein Child Health Study found in the Western Cape, South Africa. Cases in this study were defined as infants who wheezed at least once in the first year of life. Controls were infants who had not wheezed at all during this time frame. We collected faecal samples from the infants at birth, six weeks, six months and 12 months. We extracted nucleic acid from all the faecal samples using the manual ZR Fecal DNA MiniPrep™ Kit. This was followed by the sequencing of the V4 region of the 16S rRNA gene using the Illumina MiSeq technique. Bioinformatic quality control checks were performed on the sequencing data. Thereafter, the contaminating genera were removed and statistical analysis was carried out. Relative abundance curves and bar plots were generated for the infant samples. Additionally, we plotted log ratio bi-plots to ascertain differences between the faecal bacterial profiles of infants in relation to the frequency of wheezing during the first year of life.

Results: At birth, 159 infants provided faecal samples. The remaining infant samples were collected as follows: 114 at six weeks, 141 at six months and 98 samples at 12 months of age. For the infants, 140 cases and age-matched controls were included in this study. The infant bacterial profile varied during the first year of life, becoming more stable over time. By six weeks of age, 16% of the infants had wheezed once and 8% had wheezed twice or more times. There was no clustering observed between the frequency of wheezing and the associated

bacterial profiles at the different ages. The results of this study show that the composition of *Lactobacillales* in the infant stool were associated with the development of recurrent wheezing. **Conclusion:** The passed and aspirate samples were found to be very similar. Significant differences between the wheezing outcomes were observed in the infant stool composition of *Lactobacillales,* though they were due to outliers. The results of this study do not show an association between wheezing and the FLVR (*Faecalibacterium*, *Lachnospira*, *Veillonella* and *Rothia*) bacteria.

3.2. Introduction

Wheezing illness is a high-pitched whistling sound made while breathing (Kusel *et al.*, 2007). It affects at least 50% of children under the age of 5 years, with about 20% of the infants wheezing twice or more times in the first 12 months of life (Martinez *et al.*, 1995; Nriagu *et al.*, 1999; Sunyer *et al.*, 2001; Smuts, Workman and Zar, 2011). Recurrent wheezing has varying inconsistent definitions (Remes *et al.*, 2001; Ly *et al.*, 2006; Blanken *et al.*, 2013). In this study, we will refer to it as two or more wheezing episodes within 12 months, as it is not prevalent in the first year of life (Ly *et al.*, 2006). It is worth noting that approximately 30% of preschool age children with recurrent wheezing will have asthma at six years of age (Taussig *et al.*, 2003). Panettieri and colleagues observed that children who wheezed recurrently in the first year of life presented with decreased lung function and increased respiratory infections (Panettieri *et al.*, 2008). A handful of researchers across the world have investigated the association between infant gut microbiota and wheezing illness in children (Murray *et al.*, 2005; Vael *et al.*, 2008, 2011; Gosalbes *et al.*, 2013; Arrieta *et al.*, 2015; Stiemsma *et al.*, 2016). However, in Africa, only one study according to our knowledge has investigated the association between infant gut bacterial profiles and wheezing illness (Amberbir *et al.*, 2011).

Two recent publications from the CHILD Study have provided substantial evidence of the involvement of four main faecal bacterial genera in the development of wheezing in infants (Arrieta *et al.*, 2015; Stiemsma *et al.*, 2016). Arrieta and colleagues discovered that the FLVR bacteria were significantly reduced at three months of age and not at one year, in the atopic wheeze group (Arrieta *et al.*, 2015). This discovery led the researchers to come to the conclusion that the FLVR bacteria offer a protective effect in infants against the development of wheezing (Arrieta *et al.*, 2015). In this same study they were able to demonstrate the causal relationship between wheezing and the FLVR (*Faecalibacterium, Lachnospira, Veillonella*

and *Rothia*) bacteria, with the use of a murine model of airway inflammation using humanized bacteria (Arrieta *et al.*, 2015). The follow-up study by Stiemsma and colleagues showed that atopic wheezing infants had reduced amounts of *Lachnospira* and increased amounts of *C. neonatale* (Stiemsma *et al.*, 2016). The results from these two studies probe the idea that only a fraction of the stool bacteria are responsible for the development of wheezing, and not the entire stool bacterial composition.

Most of the studies investigating wheezing illness and the infant bacteria have been in high income countries, used culture and targeted experimental techniques to identify the infant stool microbial community and approximately 33 % of them consisted of small sample sizes, ranging from 18 to 76 participants (Kalliomäki *et al.*, 2001; Bisgaard *et al.*, 2011; Akay *et al.*, 2014). Therefore, we have carried out this study within a prospective birth cohort in a low socio-economic community in South Africa. The objectives of this study were to describe the major components of the faecal bacterial profile during the first year of life of infants with and without wheezing, determine the risk factors influencing that composition and to investigate the association between the stool bacteria and the development of recurrent wheezing in infants.

3.3. Methods

3.3.1. Ethics Statement

Both the Drakenstein Child Health Study (DCHS) (HREC: 401/2009) and this study (HREC: 748/2015) received approval from the University of Cape Town (South Africa) Human Research Ethics Committee (HREC). Approval letter of this study is included in Appendix D.

3.3.2. Study Setting and Design

This longitudinal study was nested within a birth cohort study, referred to as the DCHS (Zar et al 2015). The DCHS is located in the Drakenstein sub-district, 60 kilometres outside of Cape Town, South Africa (Figure 3.1). The DCHS study has enrolled about 1000 mother/infant pairs and followed the children from birth up to five years of age. The main aims of the DCHS include the investigation of aetiology, progression and risk factors for childhood pneumonia. This population is stable which makes it ideal for the longitudinal study, and they have access to child health care and antenatal services (Zar *et al.*, 2015). Pregnant mothers were enrolled at either Paarl Hospital or other neighbouring primary health care facilities; Mbekweni and TC Newman clinics. Participants included in this study required informed consent from the mothers and confirmation that they were residents of the area and would not be moving during the study period. Infants were enrolled at birth and subsequent samples were collected from six weeks to 12 months of age.

Figure 3.1. Geographical location (in red) of the Drakenstein Child Health Study, Western Cape, South Africa. The red in the insert depicts the Western Cape Province, while the red in the main figure outlines the Paarl sub-district.
Source for country map:http://en.wikipedia.org/wiki/Administrative_divisions_of_South_Africa#/media/File:Map_of_South_Africa_with_English_labels.svg Source for province map: http://commons.wikimedia.org/wiki/File:Map_of_the_Western_Cape_with_municipalities_labelled_(2006).svg

3.3.3. Selection of the Study Population

Pregnant women were enrolled from Mbekweni and TC Newman clinics between 20-28 weeks of gestation. Participants were selected for this study based on the availability of infant samples collected at birth, 6 weeks, 6 months and 12 months. We defined a case as an infant who wheezed at least once in the first year of life. The first wheezing outcome was an infant who wheezed once in the first year of life. The recurrent wheezer was the second outcome and defined as an infant who wheezed two or more times in the first year. Samples of infants reported to have wheezed in the first 12 months were identified first, and then age-matched

controls were included thereafter. Therefore the controls were samples collected from infants who did not wheeze at any point during the first year of life. Age matching is a process whereby control infants whose samples were collected at ages as close as possible to the cases were identified and included in the dataset. The main criteria used to select infant samples from the DCHS database was an infant with a stool sample collected within the time points of interest; birth, 6 weeks (4 – 8 weeks), 6 months (5 – 7 months), 12 months (11 – 13 months). There were two stool sampling types collected from infants in this study, namely: aspirated stool and normal passed stool (gastrointestinal tract stool).

3.3.4. Collection of Faecal Samples and Metadata Variables

The first stool samples were collected at the birth of the infant at the hospital. Subsequent stool samples were either collected at home or at the clinic. If they were collected at home, they were stored at -20ºC until being transported to the clinic where they would then be sent to the Division of Medical Microbiology to be stored at -80°C. Samples collected at the clinic were transported to the Division of Medical Microbiology, University of Cape Town, on the same day and stored at -80°C. Stool samples were collected in sterile, labelled screw-cap containers with a scoop attached to the lid. At the scheduled clinic visits, stool samples were collected from the babies, as well as aspirated stool samples were collected from the infants unable to pass stool. A thin plastic tube lubricated with K-Y Jelly was attached to a syringe, inserted into the anus and used to rinse out the rectum using Molecular grade water (Sigma-Aldrich). The aspirate was dispensed into a sterile 2 ml Eppendorf tube and stored at -20°C until transportation.

Information regarding medication use (antibiotics) and hospitalization was obtained from the mothers. The presence of common risk factors for gut bacterial colonization and wheezing

illness were assessed in questionnaires that were performed by healthcare professionals at the clinic at all the recorded sample collection points (birth, 6 weeks, 6 months and 12 months). Once captured, the data was stored on a central database under the management of the data team which verified the entries with the hard copy questionnaires. The questionnaires included the following variables: age, gender, mode of delivery, gestational age, current breastfeeding, introduction of solids, daycare attendance (assessed at each time point), maternal education, maternal employment status and average household income, maternal HIV status, indoor air pollution, number of people living in the house and sharing the bedroom with the baby, parental smoking as well as pet exposure.

3.3.5. Bacterial DNA Extraction from Faecal Samples

DNA was extracted from approximately 50mg of passed stool or 100µl of aspirated stool using the ZR Fecal DNA MiniPrep™ (Catalog No. D6010, Zymo Research, CA, USA), with minor changes. We transferred a maximum of 600µl of the supernatant to the Zymo-Spin™ IV Spin Filter so as to maximise the amount of DNA being processed.

3.3.6. DNA Concentration and DNA Purity

The bacterial DNA concentration and purity were determined to assess the quality of the samples before library preparation for NGS. The bacterial DNA concentration was measured by using a quantitative PCR, according to the protocol by Bogaert (Bogaert *et al.*, 2011). The purity of the DNA was assessed using the NanoDrop spectrophotometer (Thermo Scientific NanoDrop 2000 Spectrophotometer, Thermo Scientific, Wilmington, DE 19810 USA). DNA samples failing to amplify were excluded from library preparation. The 260/280 and 260/230 ratios from the nanodrop were used to assess the purity of the nucleic acid. The best 260/280 ratios range from 1.8 to 2.0, and anything outside this range indicates the presence of

contaminants such as phenols and proteins that absorb at 280nm. The optimum 260/230 ratios range from 2.0 to 2.2, and anything outside this range indicates the presence of contaminants that absorb at 230nm.

3.3.7. 16S ribosomal RNA gene Library Preparation

Figure 3.2 summarizes the sequence of steps for library preparation of the 16S rRNA gene for the samples under investigation.

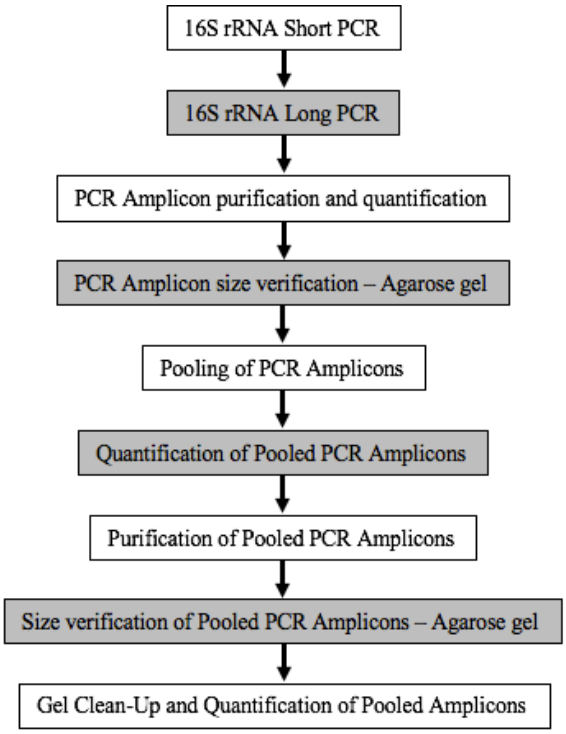


Figure 3.2. Flow diagram representing the various steps involved in the 16S rRNA Library Preparation.

3.3.7.1. PCR Amplification of 16S ribosomal Bacterial DNA

Two PCR reactions were carried out to amplify and add barcodes to the V4 hypervariable region of the 16S rRNA gene, with minor modifications (Caporaso *et al.*, 2011).

3.3.7.2. 16S rRNA Short PCR

The first PCR was performed in 25.25µl and contained 1X MyTaq™ HS Mix (Bioline, MA, USA), 0.8µM of each modified primer F515 short: (5' GTGCCAGCHGCYGCGGT 3') and R806 short: (5' GGACTACNNGGGTWTCTAAT 3'), 2.97% dimethyl sulfoxide (Sigma-Aldrich®, MO, USA) and 4µl of template DNA per sample. Cycling conditions of the first PCR included an initial denaturation step at 95ºC for 3 min, 10 cycles of denaturation at 95ºC for 30 sec, annealing at 50ºC for 30 sec and extension at 72ºC for 1 sec; followed by a final extension step at 72ºC for 5 min. Each PCR run included a positive (25 ng Cyanobacteria genomic DNA) control, non-template controls and technical repeats with 4µl of template DNA. PCR reactions were performed in an Applied Biosystems® 2720 Thermal Cycler (CA, USA).

3.3.7.3. 16S rRNA Long PCR

The second PCR reaction included the primers F515-composite (5′-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNNNNNNGTGCCAGCHGCYGCGGT-3′) and R806-composite (5′CAAGCAGAAGACGGCATACGAGATACGAGACTGATTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNGGACTACNNGGGTWTCTAAT-3′) (Caporaso *et al.*, 2011). This PCR reaction of 25.25µl contained 1X MyTaq™ HS Mix (Bioline, MA, USA), 0.8µM of each primer, 2.97% dimethyl sulfoxide (Sigma-Aldrich®, MO, USA) and 4µl of amplicon from the first PCR reaction. In the second PCR, the amplicon from the first PCR was used to attach the adapters, and this reaction also included barcodes, priming region and 12-15 staggered nucleotides (NNNNNNNNNNNN). The cycling conditions were identical to the first PCR, except that the PCR included 30 cycles.

3.3.7.4. PCR Amplicon purification and quantification

PCR amplicons from the second round were purified using Agencourt AMPure XP (Beckman Coulter, CA, USA), following the manufacturer's guidelines with one modification. The modification applied in this study entailed dispensing 16.25µl (0.65 x 25µl reaction volume per well) beads into each well and pipette mixing. We lastly transferred 35µl of the eluent to a new plate and stored at 4°C.

The quantification of the purified PCR product was done using the QuantiFluor® dsDNA System in 96 well plates according to the manufacturer's instructions. The GloMax®-Multi Microplate Multimode Reader (WI, USA) was used to quantify each of the purified amplicons.

3.3.7.5. PCR Amplicon size verification

Aliquots of purified PCR amplicons (5µl) were loaded onto 2% agarose gels containing 0.1µg/ml of Ethidium Bromide. TrackIt™ 1 Kb Plus DNA Ladder (Thermo Fischer, USA) (5µl) was loaded on all agarose gels. Agarose gels were electrophoresed for 90 min at 110 Volts. Agarose gels were viewed using the Chemi Genius 2 Bio Imaging System (CA, UK).

3.3.7.6. Pooling of PCR Amplicons

Based on the amplicon concentrations measured above, the required volume for each purified amplicon was calculated in order to contribute towards the 80ng of pooled DNA. The 96 well plates were spun before transferring the estimated amplicon amounts from each well to a 1.5ml Eppendorf tube.

3.3.7.7. Quantification of Pooled PCR Amplicons

The amplicon pooled DNA was quantified using the Qubit® dsDNA BR Assay Kit (Invitrogen™, Eugene, OR, USA) according to the manufacturer's recommendations. This kit

was used because it measures double-stranded DNA in the concentration range from 100pg/µl to 1000ng/µl.

3.3.7.8. Purification of Pooled PCR Amplicons

Pooled DNA was purified using the Agencourt AMPure XP (Beckman Coulter, CA, USA) following the manufacturer's guidelines with the following modification; beads were dispensed in a 1:1 ratio. We transferred 60µl of the eluent to a new tube and stored at 4°C. The purified pooled DNA was quantified as above.

3.3.7.9. Size verification of Pooled PCR Amplicons

A 1.5% agarose gel containing Ethidium Bromide (0.1µg/ml) was prepared. Three wells of the comb were taped together, to create a large well that could hold more volume of sample. This large well allowed 6µg of amplicon pooled DNA with loading dye to be loaded into it. TrackIt™ 1 Kb Plus DNA Ladder (Thermo Fischer, USA) was included on the agarose gel. In addition, 20µl of Ethidium Bromide was added to the Tris-Acetate EDTA buffer tank at the anode end (at the red electrode). Gel electrophoresis conditions were set as follows: 35 volts for 30 min, then 40 volts for 45 min and finally 70 volts for 3 hrs 30 min. Agarose gels were viewed on a long wavelength (312nm) FOTODYNE Transilluminator (WI, USA) to excise DNA ranging between 400bp and 500bp from the gel, using a sterile blade. The gel slice was trimmed to remove excess gel before placing the DNA-containing gel slice in a sterile 1.5ml Eppendorf tube and viewing it using the Chemi Genius 2 Bio Imaging System (CA, UK).

DNA was purified from the agarose gel slices using the QIAquick Gel Extraction Kit (USA), according to the manufacturer's guidelines.

3.3.7.10. Quantification of Pooled Amplicons

Gel extracted DNA was quantified using the Qubit® dsDNA BR Assay Kit (Invitrogen™, Eugene, OR, USA) according to the manufacturer's recommendations.

3.3.8. Sequencing of 16S ribosomal RNA V4 region

Faecal samples collected from the infants, as well as a set of controls were sequenced in two sequencing runs by the Centre for Proteomic and Genomic Research (CPGR), Cape Town, South Africa. Sequencing run 1 contained 245 samples and 24 sequencing controls. Sequencing run 2 contained 267 samples and 18 sequencing controls. The sequencing controls for each run consisted of two positive controls; genomic DNA from a mock microbial community known as a BEI control (HM-783D, BEI Resources, VA, USA) and genomic DNA from Cyanobacteria (Arthrospira). The other controls included were non-template controls which consisted of no-template DNA extraction controls and Cyanobacteria spiked no-template DNA extraction controls, as well as technical repeats which consisted of nucleic acid randomly selected from the samples.

CPGR conducted a quality control check which entailed the quantitation of adapter-ligated double-stranded DNA (dsDNA) on the 7900HT Fast Real-Time PCR System using the KAPA Library Quantification Kits (Illumina®), and analysis of fragment size distribution on the Agilent 2100 Bioanalyzer using the Agilent High-Sensitivity (HS) DNA Kit. Thereafter, the library pool was diluted to 4 nM using Buffer EB (Qiagen, DE). A 5 µl aliquot of the 4 nM library was denatured by adding 5 µl of 0.2 N NaOH. The mixture was incubated at room temperature for 5 min, and 990 µl of HT Buffer (Illumina®) was added to bring the library concentration to 20 pM. The library was further diluted to 4 pM with HT Buffer. The PhiX

library was diluted from 10 nM to 4 nM with Buffer EB and then denatured as above and diluted to 4 pM. The PhiX library was spiked into the library at 15% (v/v) and further denatured at 96 °C for 2 min and incubated on ice for 5 min before loading onto the MiSeq® instrument for sequencing using the MiSeq® Reagent v3 kit (600 cycle). The MiSeq® system was programmed for a 2x 251-cycle, paired-end sequencing run to generate FASTQ files.

3.3.9. Bioinformatic and Statistical Analysis

3.3.9.1. Bioinformatic Analysis

The sequencing quality of FASTQ files was analyzed using the Fastqc package (Wang *et al.*, 2007; Andrews, 2010; Edgar, 2010, 2013). UPARSE and USEARCH were used to assess, sort, carry out the dereplication, clustering and chimera removal from the sequences as described by Pylro and colleagues (Wang *et al.*, 2007; Edgar, 2010, 2013; Pylro *et al.*, 2014). A few modifications to the above protocol included the removal of contamination as detailed below, and the use of SILVA database for taxonomy assignment. The faecal sample sequences with a matching identity of 100% to no-template control sequences were removed from the faecal samples if they had the same or lower number of reads compared to the average reads of the contaminant Operational Taxonomic Units (OTUs) in the controls. Otherwise, the average number of reads of the contaminants were subtracted from the matching biological sequences. Further processing of the data was carried out using the Quantitative Insights Into Microbial Ecology (QIIME 1.7.0) suite of software tools (Caporaso *et al.*, 2010). The taxonomy was therefore assigned as mentioned by Claassen-Weitz and colleagues (Claassen-Weitz *et al.*, 2018). Rarefaction curves were calculated for all the samples included in this study.

3.3.9.2. Statistical Analysis

The software applications used in this study for statistical analysis and graphical data representations were R 3.4.1 and RStudio 1.1.414. The output from the Illumina MiSeq

sequencing software QIIME was used for the downstream statistical analysis. The R packages used for the analysis included biomformat, car, cluster, MASS, psych, smacof and vegan (Venables and Ripley, 2002; de Leeuw and Mair, 2009; Fox and Weisberg, 2011; Hastie, 2013; Oksanen, Blanchet and Kindt, R, 2013; Maechler *et al.*, 2015; McMurdie and Paulson, 2016; Liaw and Wiener, 2018; Revelle, 2018). Box plots were drawn to graphically show the statistically significant different medians at 5% significance ($p<0.05$).

3.3.9.3. Analysis of the Sociodemographic Data

R package epicalc was used to describe the characteristics of the participants, with the use of summ and tab1 functions. The clinical data in conjunction with the relative abundance data was analysed in R by use of the vegan package and generalised linear models. Anova tests were also computed as part of the analysis of variance. A p-value less than or equal to 0.05 was considered statistically significant.

3.3.9.4. Calculating beta diversity of aspirate stool samples, passed stool samples and spiked aspirate sequencing controls

Beta diversity is a measure of inter-individual diversity which is calculated between different samples (Whittaker, 1960). For this reason, the similarity between the aspirate stool and the passed stool samples were measured using beta diversity. The first measures used was β_w: (Whittaker, 1960), which ignores the relative magnitude of the species found in one sample but not the other. The second measure was β_{sim}: (Lennon, J.J., Koleff, P., Greenwood, J.J.D., Gaston, 2001) based on (Simpson, 1943), which focuses on compositional difference, more than difference in species richness. Beta diversities were computed pairwise, with $\beta = 0$ indicating perfect similarity and $\beta = 1$ indicating complete difference. Here we computed the mean beta diversity between the aspirate stool samples, passed stool samples and the spiked (using *Cyanobacteria*) aspirate sequencing controls as shown in Table 3.1.

Table 3.1. Table showing how the mean beta diversities were computed between the samples.

	Sample $X1$	Sample $X2$	...	Sample Xn
Sample $X1$				
Sample $X2$				
...				
Sample Xn				

The number of 'pairwise' comparisons is n^2, with the values in the upper triangle $\left((n-1)+(n-2)+\cdots+1=\frac{n(n-1)}{2}\right)$ equal to those in the lower triangle since $\beta(X_i,X_j)=\beta(X_j,X_i)$. The values on the diagonal are all zero since it is a measure of the beta diversity between a sample and itself. These zero values are included in the computation of the mean of all n^2 values. The total of all beta diversity values in the matrix is:

$$n^2\bar{B}=\sum_{i=1}^{n}\sum_{j=1}^{n}\beta_{ij}=\sum_{i\neq j}\beta_{ij}$$

The mean beta diversity of the off-diagonal values is obtained by adjusting the total for division by the total number of off-diagonal elements $n(n-1)$ to obtain:

$$\frac{1}{n(n-1)}n^2\bar{B}=\frac{n}{n-1}\bar{B}$$

3.3.9.5. Multidimensional scaling of clusters

Multidimensional scaling was executed in order to determine the similarity between the aspirated and the passed stool samples included in this study. The smacof R package was used to generate the cluster plots (de Leeuw and Mair, 2009).

3.3.9.6. Summary Statistics

Box and whisker plots were drawn to show the distribution of the DNA concentrations and purities of all the stool samples.

3.3.9.7. Evaluation of sequencing depth

Accumulation or rarefaction curves were constructed to determine if adequate numbers of reads were obtained from all the stool samples of infants, to identify all bacterial genera (Hughes *et al.*, 2016). The sequencing efficiency was assessed by constructing accumulation curves of the number of genera against the number of reads obtained (richness), using the function [rarefy] from the R package vegan (Hurlbert, 1971; Oksanen, Blanchet and Kindt, R, 2013).

3.3.9.8. Calculation of the Species Richness (Alpha Diversity)

The Shannon diversity index (H') was calculated to determine the alpha diversity of the infants included in this study (Shannon, 1948; Morgan and Huttenhower, 2012). Alpha diversity measures the evenness and richness of bacteria in the sample of interest. A larger Shannon diversity index indicates greater diversity within the sample.

$$H' = \sum_{i=1}^{s} (p_i)(\ln p_i)$$

Where i = species of interest,
pi = proportion of species of interest (i) relative to the total number of species,
ln = natural logarithm,
Σ = sum of the resulting product across species, and
s = number of species recorded.

3.3.9.9. Generation of log ratio biplots using the sequencing data

Log ratio biplots simultaneously plot compositional data on a multidimensional platform (McMurdie and Paulson, 2016). We generated log ratio biplots using the compositional data of

the samples in the OTUs that were shown to be statistically different, at 5% significance (Greenacre, 2010).

3.3.9.10. Analysis of wheezing relationship in infants

Log ratio bi-plots were generated based on wheezing frequency in the first year of life at both phylum and genus level (McMurdie and Paulson, 2016). Coplots of bacterial taxa and wheezing variables which included wheeze onset and number of wheezing episodes in the first year of life were also generated as part of the analysis of the wheezing relationship in infants.

3.4. Results

3.4.1. Participant Characteristics

3.4.1.1. Sample Collection and Distribution

During the study design, we pre-determined that a subset of 140 cases and 140 age-matched controls would be used in the investigation for this nested study. A total of 512 stool samples were collected from 280 infants between August 2012 and October 2015 (Figure 3.3). At birth, 159 infants provided stool samples, in the form of passed stool samples. The remaining infant stool samples were collected as follows: 114 collected at six weeks, 141 collected at six months and 98 stool samples collected at twelve months of age (Figure 3.3). From six weeks to 12 months the number of aspirated stool samples ranged from 9 – 33% (Table 3.2). The heat map (Figure 3.4) shows the sample collection for the infants at the time points under investigation. The heat map also indicates that the majority of the infants did not have stool samples collected at all the times points. For the cases, 46 had one sample, 49 had two samples, 20 had three samples and 14 had four samples collected. For the controls, 45 had one sample, 44 had two samples, 21 had three samples and 14 had four samples collected.

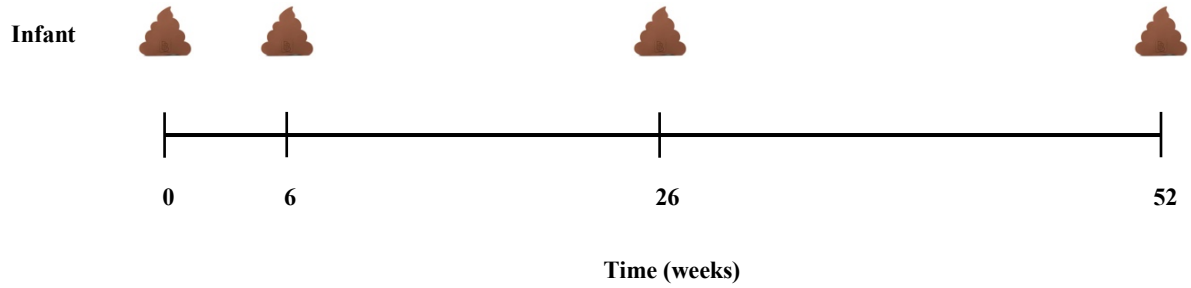


Figure 3.3. Schematic of the longitudinal data collection. Samples were collected at birth from the infants (159), and the infants were followed up at 6 weeks (114), 6 months (141) and 12 months (98) for subsequent stool sample collection.

Table 3.2. Sample distribution between the two types, aspirated stool and passed stool.

Age	Aspirate stool (%)	Passed stool (%)	Total
birth	0 (0)	159 (100)	159
6 weeks	10 (9)	104 (91)	114
6 months	21(15)	120 (85)	141
12 months	32 (33)	66 (67)	98

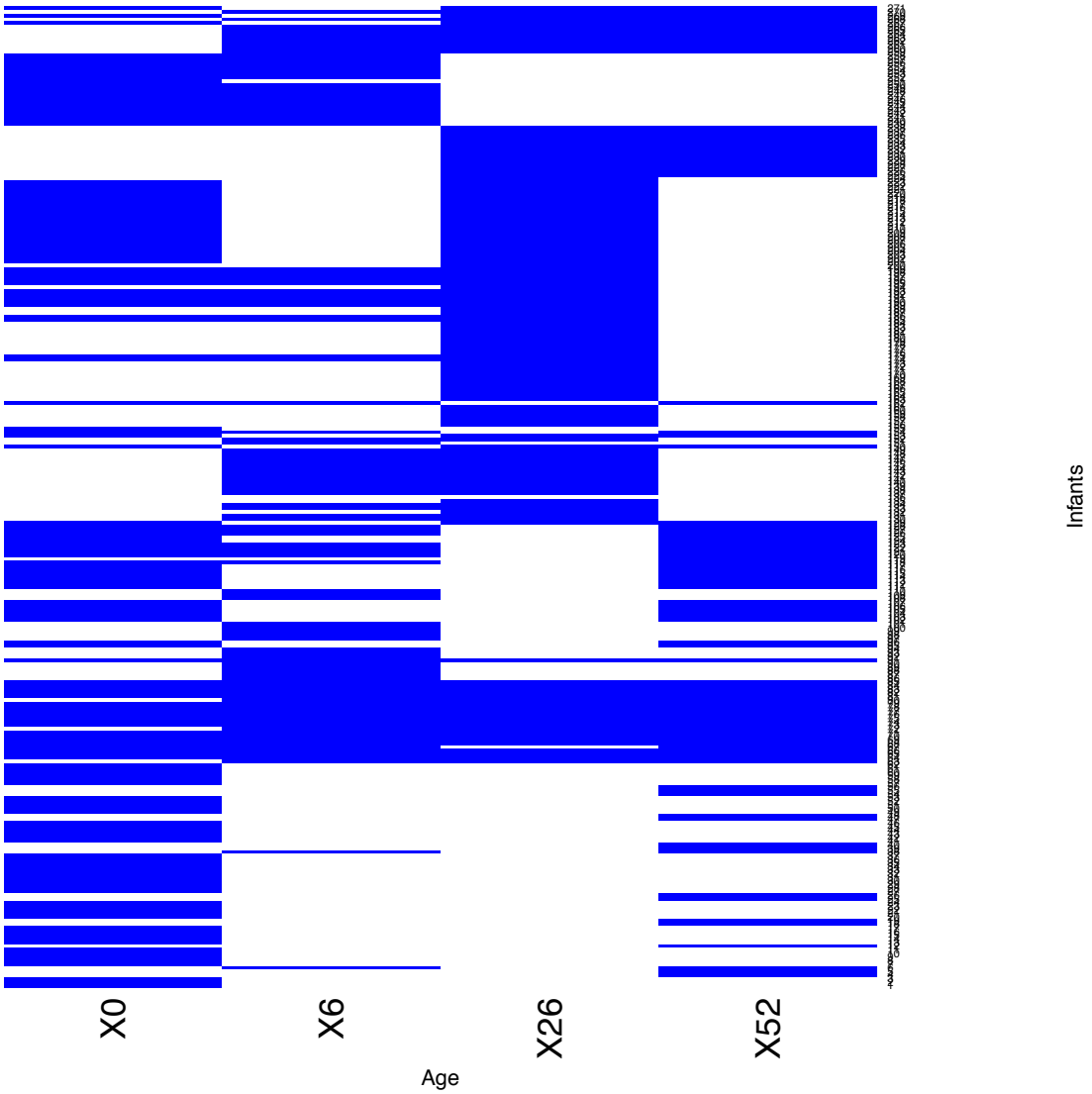


Figure 3.4. Heat map of sample collection of the infants from birth to 12 months. Each row represents an infant, and the different columns represent the different time points namely birth, six weeks, 26 weeks and 52 weeks.

3.4.1.2. Study Population

Table 3.3 summarizes the study participants clinical and demographic characteristics as split between cases and controls. The median age of the 159 infants that provided a faecal sample at birth was one day (25-75 percentile (Q1-Q3; 0 – 1)). The 114 infants that provided a sample at six weeks had a median age of seven weeks (Q1-Q3; 6.5 – 8.5), and the 141 infants with a six-month sample had a median age of six months (Q1-Q3; 5.8 – 6.2). Lastly, 98 infants provided a faecal sample at 12 months, and the median age was 12 months (Q1-Q3; 11.7 – 12.5). Over 90% of the infants included in this study were full term with a median birth weight

of 3.2 kg (Q1-Q3; 2.7 – 3.5), and median birth length of 50 cm (Q1-Q3; 48 – 52). Most of the infants (~80 %) were vaginally delivered. Infant smoke exposure from the mother was 57% in the cases and ~63% in the controls. The majority of the mothers in both groups (>95%) received secondary or tertiary education.

Table 3.3. Summary of participant characteristics.

Characteristics	Cases	Controls
Number	140	140
Residential Area:		
Mbekweni, n (%)	55 (39.3%)	53 (37.9%)
TC Newman, n (%)	85 (60.7%)	87 (62.1%)
Maternal education:		
Primary level, n (%)	6 (4.3%)	3 (2.1%)
Secondary level, n (%)	126 (90%)	124 (88.6%)
Tertiary level, n (%)	8 (5.7%)	13 (9.3%)
Infant HIV status:		
HIV-infected, n (%)	2 (1.4%)	-
HIV-uninfected, n (%)	138 (98.6%)	140 (100%)
Maternal smoking status:		
Smoker, n (%)	80 (57.1%)	88 (62.9%)
Non-Smoker, n (%)	60 (42.9%)	52 (37.0%)
Mode of delivery:		
Vaginal delivery, n (%)	113 (80.7%)	114 (81.4%)
Caesarean-section delivery, n (%)	27 (19.3%)	26 (18.6%)
Gestational age (weeks), median (Q1-Q3) [§]	39 (38-40)	39 (38-40)
Birth weight (kg) , median (Q1-Q3) [§]	3.0 (2.7-3.4)	3.2 (2.9-3.5)
Birth length (cm) , median (Q1-Q3) [§]	50 (48-52)	50 (49-52)
Gender:		
Male, n (%)	76 (54.3%)	76 (54.3%)
Household members, n (Q1-Q3) [§]	5 (3-6)	5 (3-7)

[§] - Interquartile range

3.4.1.3. Wheezing Data

The number of infants experiencing one wheezing episode during the study period increases over time from six weeks to 12 months (Table 3.3). At the six-week follow-up, 16% of the infants had wheezed once. A further 9% (13/140) of the infants wheezed two or more times in the same time frame. At six months 21% of infants wheezed once and a further 31% wheezed two or more times. At 12 months, 29% of the infants wheezed once and 16% wheezed twice or more, from the age of six months. In this dataset, 40.8% of the infants wheezed once, and 59,7% wheezed more than once. The heat map shows the distribution of the wheezing in the infants over time (Figure 3.5). It also shows that at age 6 months there was the largest proportion of infants with more than one wheezing episode in this dataset.

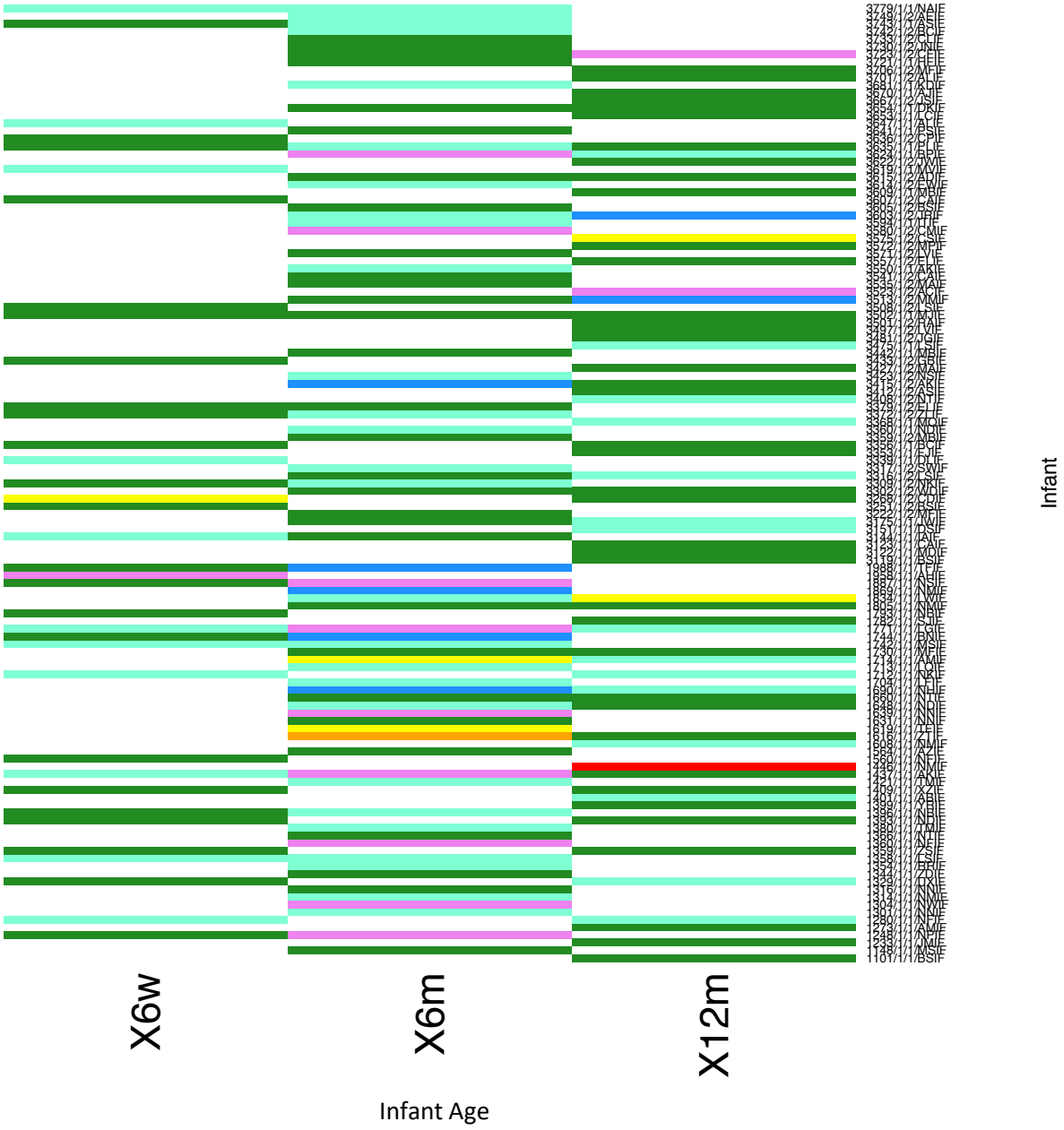


Figure 3.5. Heat map of infant wheezing data.
Each row represents an infant, and the different columns represent the different time points namely; six weeks, six months and 12 months. The different colours represent the number of times the infant has wheezed in total up to that time point. A blank in the heatmap either means that the infant did not wheeze at that timepoint or the data was not available. Forestgreen – 1 wheeze episode, Aquamarine – 2 wheeze episodes, Violet – 3 wheeze episodes, Dodgerblue - 4 wheeze episodes, Yellow - 5 wheeze episodes, Orange - 6 wheeze episodes, Red - 7 wheeze episodes.

3.4.2. Quality Control of the DNA and Sequencing Data

3.4.2.1. Determination of the Concentration and Purity of the DNA

The baby's first stool (meconium) sample median DNA concentration was 2.9 ng/µl (Q1-Q3; 1.7 – 6.2) (Figure 3.6A). The median DNA concentration of the passed stool samples was

318.75 ng/µl (Q1-Q3; 175.92 – 549.25) (Figure 3.6A). The median DNA concentrations of the aspirate stool samples were significantly lower than the passed stool samples (19.44 ng/µl (Q1-Q3; 6.09 – 57.98) vs 318.75 ng/µl (Q1-Q3; 175.92 – 549.25); p= 8.27E-11). The median purity of the passed stool samples as measured by the Nanodrop 260/280 ratio was 1.63 (Q1-Q3; 1.16 – 1.70) (Figure 3.6B). The median purity of the passed stool samples as measured by the Nanodrop 260/230 ratio was 1.10 (Q1-Q3; 0.43 – 1.30) (Figure 3.6C). The median purity of the aspirate stool samples as measured by the Nanodrop 260/280 ratio was 1.35 (Q1-Q3; 1.22 – 1.48) (Figure 3.6B). The median purity of the aspirate stool samples as measured by the Nanodrop 260/230 ratio was 0.57 (Q1-Q3; 0.52 – 0.65) (Figure 3.6C).

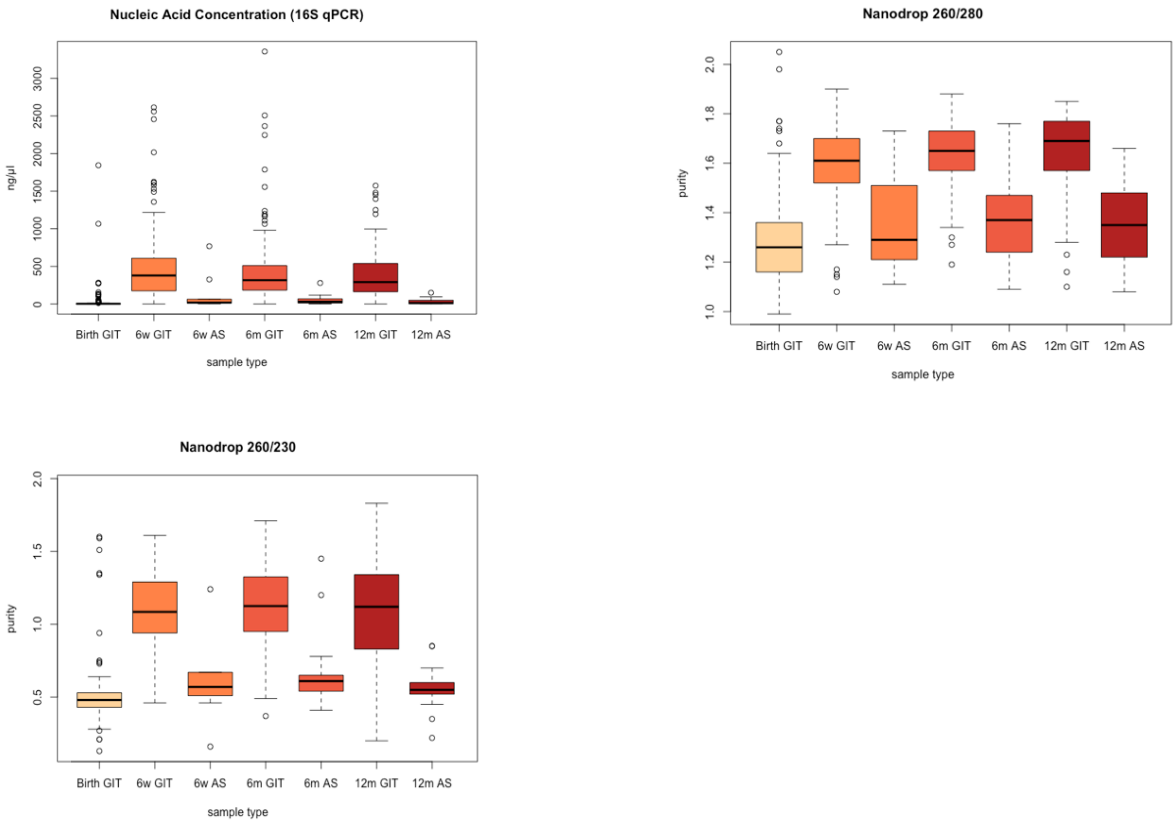


Figure 3.6. Qualitative and quantitative analysis of the DNA.
(A) Concentration of the DNA as determined by 16S qPCR **(B)** Purity of the DNA (260/280 ratio) as determined by the Nanodrop spectrophotometer **(C)** Purity of the DNA (260/230 ratio) as determined by the Nanodrop spectrophotometer.

3.4.2.2. Sequencing of Pooled 16S ribosomal RNA gene Libraries

The concentrations of adapter-ligated library in the pooled library samples were 35.5 nM and 29.75 nM, for Runs 1 and 2 respectively, as determined by the KAPA Library Quantification Kit. The average fragment sizes for Run 1 and Run 2 were 459 bp and 448 bp respectively, as determined by the Bioanalyzer (Figure 3.7). The cluster density for Run 1 was 502 k/mm^2, and a total of 5.7 Gb of data was obtained. Of the reads mapped to the indices used, 74.8 % passed the filter set by the on-board sequencing software, and 3.7 Gb (64.9%) had a quality score of Q30 and above (Figure 3.8). The cluster density for Run 2 was 495 k/mm^2, and a total of 6.57 Gb of data was obtained. Of the reads mapped to the indices used, 75.9 % passed the filter set by the on-board sequencing software, and 4.2 Gb (63.9%) had a quality score of Q30 and above (Figure 3.8).

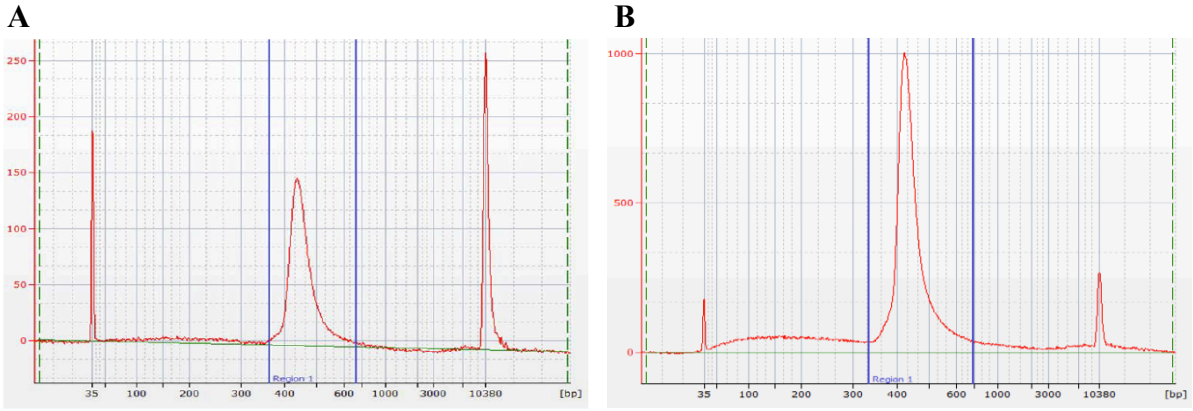


Figure 3.7. Fragment size distribution analysis of the pooled library on the Agilent 2100 Bioanalyzer using the HS DNA Kit.
(A) Run 1 1:100 dilution of the pooled library is displayed in the electropherogram **(B)** Run 2 1:10 dilution of the pooled library is displayed in the electropherogram.

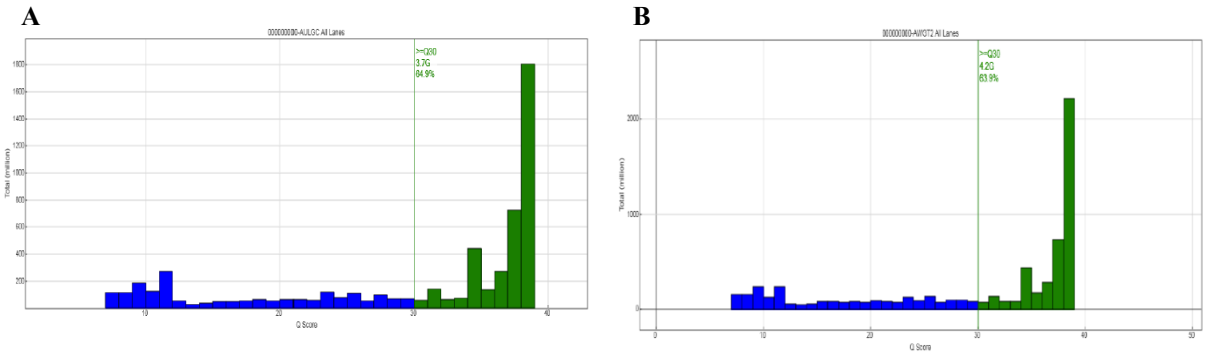


Figure 3.8. Q-Score Distribution Plot.
(A) The percentage of bases with a quality score of Q30 and above was 64.9 % (3.7Gb) for Run1, and **(B)** 63.9 % (4.2Gb) for Run2.

3.4.2.3. Bioinformatics Analysis

A total of 14 921 435 reads were generated from the two sequencing runs. A median number of reads of 18 766 (IQR; 15 362 – 24 041) was obtained per sample. The number of reads decreased by 87.8% to 1 820 540, after the initial quality control steps that included merging, quality filtering, dereplication and removal of chimeras (Table 3.4) were performed. The sequencing reads decreased by another 6.9% after the removal of contaminants. The total number of reads thus used in the analysis for the two runs was 1 694 469 (Table 3.4).

Table 3.4. The number of reads generated after the different quality control steps.

Process	Number of Reads
After Sequencing	14 921 435
Before contamination removal	1 820 540
After contamination removal	1 694 469
Used for Analysis	1 694 469

3.4.2.4. Rarefaction Curve

The rarefaction curves of the infants at birth, 6 weeks and 12 months reached the plateau phase, indicating sufficient sequencing depth for measures of diversity. However, the rarefaction curves of the infants 6 month samples had insufficient sequencing depth (Figure 3.9).

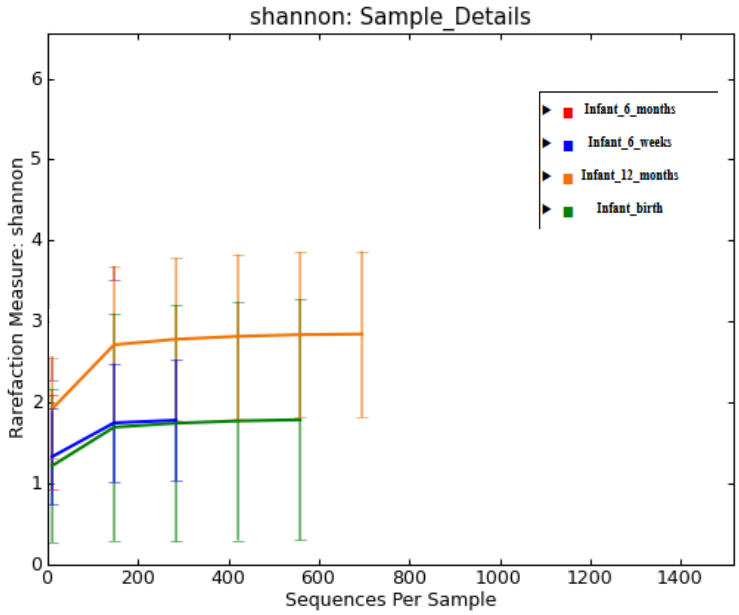


Figure 3.9. Rarefaction curves of all the samples sequenced via the V4 region of the 16S rRNA gene.

3.4.2.5. The similarity between the aspirate stool and the passed stool (beta diversity)

The aspirated samples and passed stool samples were similar to each other according to the mean beta diversities using β_w. Table 3.5 shows the mean beta diversities at the different time points, from 6 weeks to 12 months. At all the time points the aspirated samples were very different when compared to the aspirated controls as represented by the high beta diversities (Table 3.5).

Table 3.5. Mean beta diversities as calculated using β_w.

	6 weeks			6 months			12 months		
	Control	**AS**	**PS**	**Control**	**AS**	**PS**	**Control**	**AS**	**PS**
Control	0.260	0.961	0.957	0.260	0.947	0.949	0.260	0.947	0.959

AS	0.961	0.627	0.631	0.947	0.650	0.666	0.947	0.621	0.670
PS	0.957	0.631	0.571	0.949	0.666	0.678	0.959	0.670	0.615

AS – Aspirated stool. PS – Passed stool. Control – spike (*Cyanobacteria* DNA) extracted and sequenced aspirate water blank.

The mean beta diversities using β_{sim} showed that the aspirated samples and passed stool samples were similar to each other. Table 3.6 shows the mean beta diversities using β_{sim} at the different time points, from 6 weeks to 12 months. At all the time points the aspirated samples were very different when compared to the aspirated controls as shown by the high beta diversities (Table 3.6). In both cases, the pattern was similar for β_w and β_{sim}. As a result, for the downstream analysis, both aspirate and passed stool samples were included.

We observed that the infant passed stools clustered together with the aspirated stools (Figure 3.10). The two sample types overlapped, showing that they cannot be differentiated.

Table 3.6. Mean beta diversities as calculated using β_{sim}.

	6 weeks			**6 months**			**12 months**		
	Control	**AS**	**PS**	**Control**	**AS**	**PS**	**Control**	**AS**	**PS**
Control	0.151	0.955	0.930	0.151	0.913	0.910	0.151	0.918	0.932
AS	0.955	0.331	0.376	0.913	0.488	0.518	0.918	0.462	0.482
PS	0.930	0.376	0.418	0.910	0.518	0.530	0.932	0.482	0.442

AS – Aspirated stool. PS – Passed stool. Control – spike (*Cyanobacteria*) DNA extracted and sequenced aspirate water blank.

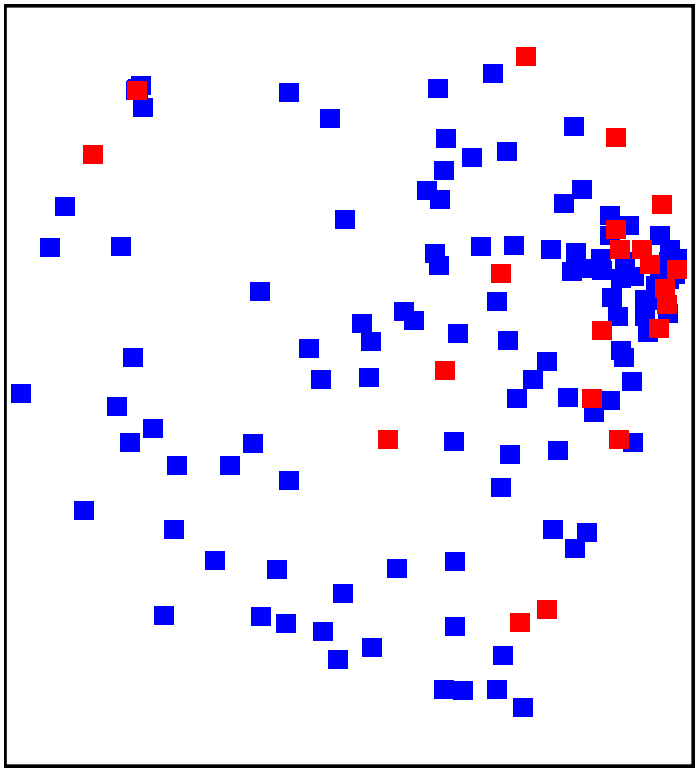


Figure 3.10. Multi-dimensional scaling (MDS) plots of the infant aspirated and passed stool at 6 months. Red – aspirated stool samples, blue – passed stool samples.

3.4.3. Faecal bacterial profile

3.4.3.1. Infant bacterial profiles in the first year of life (phylum level)

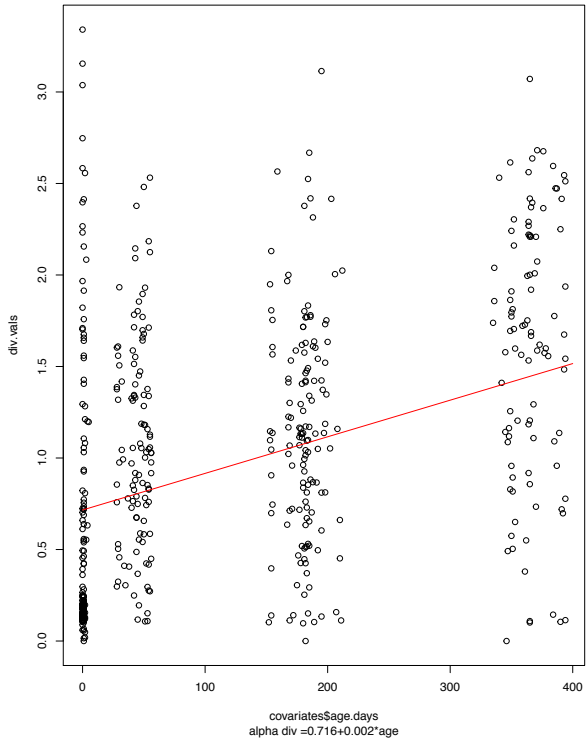


Figure 3.11. Overall diversity plot for the infant samples.

The x-axis shows the covariates and age in days. The y-axis shows the diversity values.

The overall diversity in the infant stool bacteria increased with age (Figure 3.11). Characterization of the stool bacteria of infants in this study shows that at the four time-points, there were distinct differences in the bacterial profiles (Figure 3.12). At birth, the phylum *Proteobacteria* (84.4%) was the most abundant in the infant stool samples (n=159), followed by the phyla Firmicutes (9.1%), *Actinobacteria* (6.4%) and *Bacteroidetes* (0.1%). At six weeks (n=114), the amount of *Proteobacteria* dropped to 28.9%, and the phylum *Actinobacteria* (44.7%) became the most abundant. The phylum *Firmicutes* more than doubled in abundance from birth to six weeks and continued to increase to 38.9% at 12 months. The phylum *Bacteroidetes* remained low over the 12-month period (Figure 3.12). *Actinobacteria* was low at birth and peaked at around 100 days and dropped to around 20% relative abundance at one year (Figure 3.13 A). *Bacteroidetes* remained the lowest from birth to 12 months (Figure 3.13 B). *Firmicutes* were also low at birth and steadily rose till one year (Figure 3.13 C). *Proteobacteria* was the highest at birth and dropped to over 20% relative abundance at one year (Figure 3.13 D). Additional relative abundance plots from the faecal samples and controls included in this study are included in the Appendices A-B.

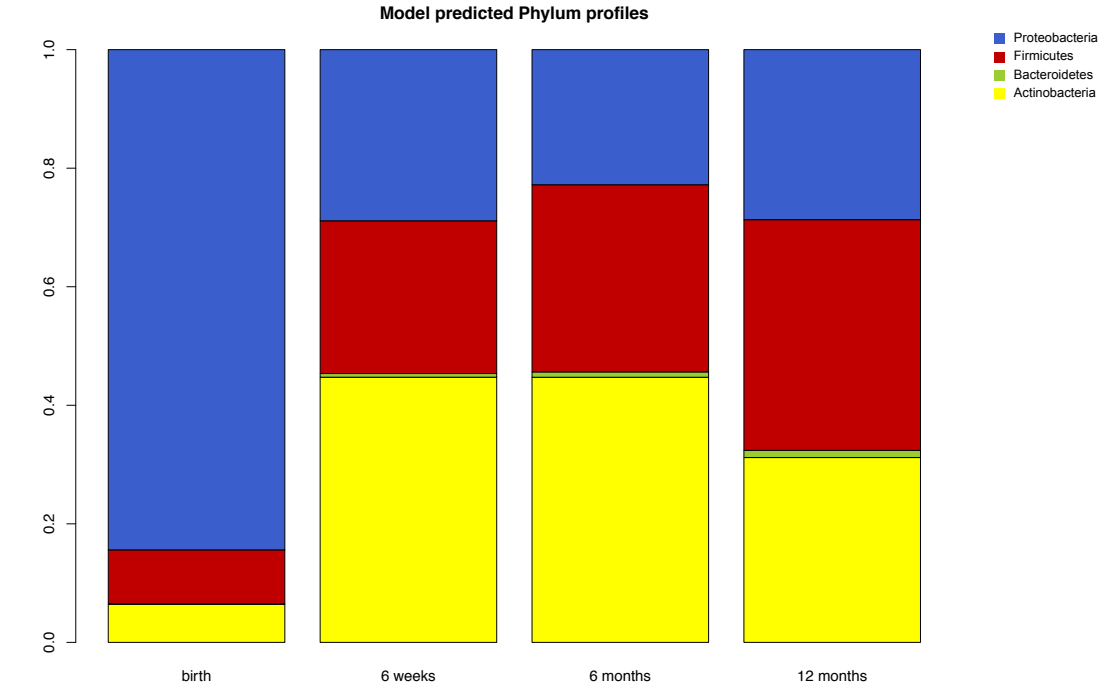


Figure 3.12. Phylum level bacterial profiles of the infant stool samples from birth to 12 months.
The relative abundances of the bacterial phyla in the infant stool samples at different time points were visualized by bar plots. Each bar represents all the subjects at that time-point, and each coloured box a bacterial taxon.

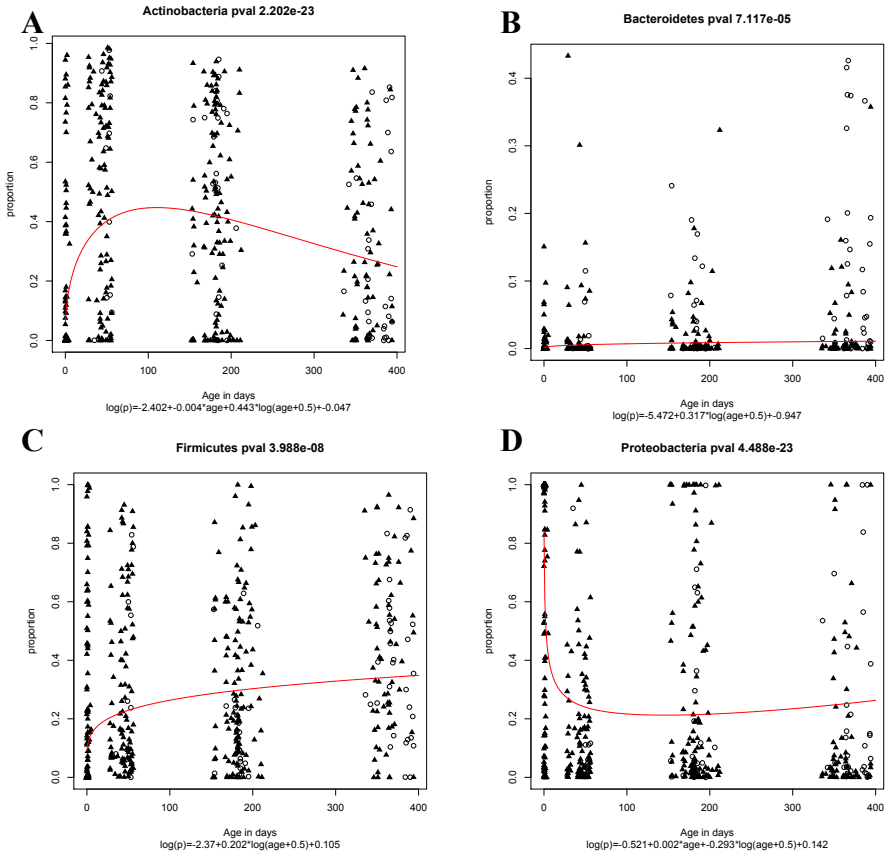


Figure 3.13. Relative abundance plots for the infants at phylum level from birth to 12 months.

(A) *Actinobacteria* **(B)** *Bacteroidetes* **(C)** *Firmicutes* **(D)** *Proteobacteria*. The x-axis shows the age in days and the y-axis shows the relative abundance of the bacteria. The red line is a plot of a model that was fitted in which age was significant.

3.4.3.2. Infant bacterial profiles in the first year of life (genus level)

At birth, the genus *Acinetobacter* (76.7%) dominated the infant stool bacteria, with the remaining genera having relative abundances below 10% (Figure 3.14). At the six weeks, six months and 12 month time points, the most abundant genus was *Bifidobacterium* at 55.3%, 55.6% and 41.2% respectively (Appendix B, Figure S8). *Escherichia-Shigella*, *Streptococcus*, *Acinetobacter* and *Blautia* were the following genera with the high relative abundances at all three time-points. At 12 months, *Bifidobacterium* (41.2%) was the most abundant genus, followed by *Acinetobacter* (16.2%), *Blautia* (9.9%), *Escherichia-Shigella* (9.2%) and *Incertae Sedis* (5.5%) (Figure 3.14). *Bifidobacterium* was low at birth and peaked at around 100 days, and dropped again at one year (Appendix B, Figure S8C). *Clostridium* remained very low from birth to 12 months (Appendix B, Figure S8E). *Faecalibacterium* was also very low at birth and steadily rose till one year (Appendix B, Figure S8I). *Lactobacillus* was low at birth, peaked around 6 week and dropped at one year (Appendix B, Figure S8K).

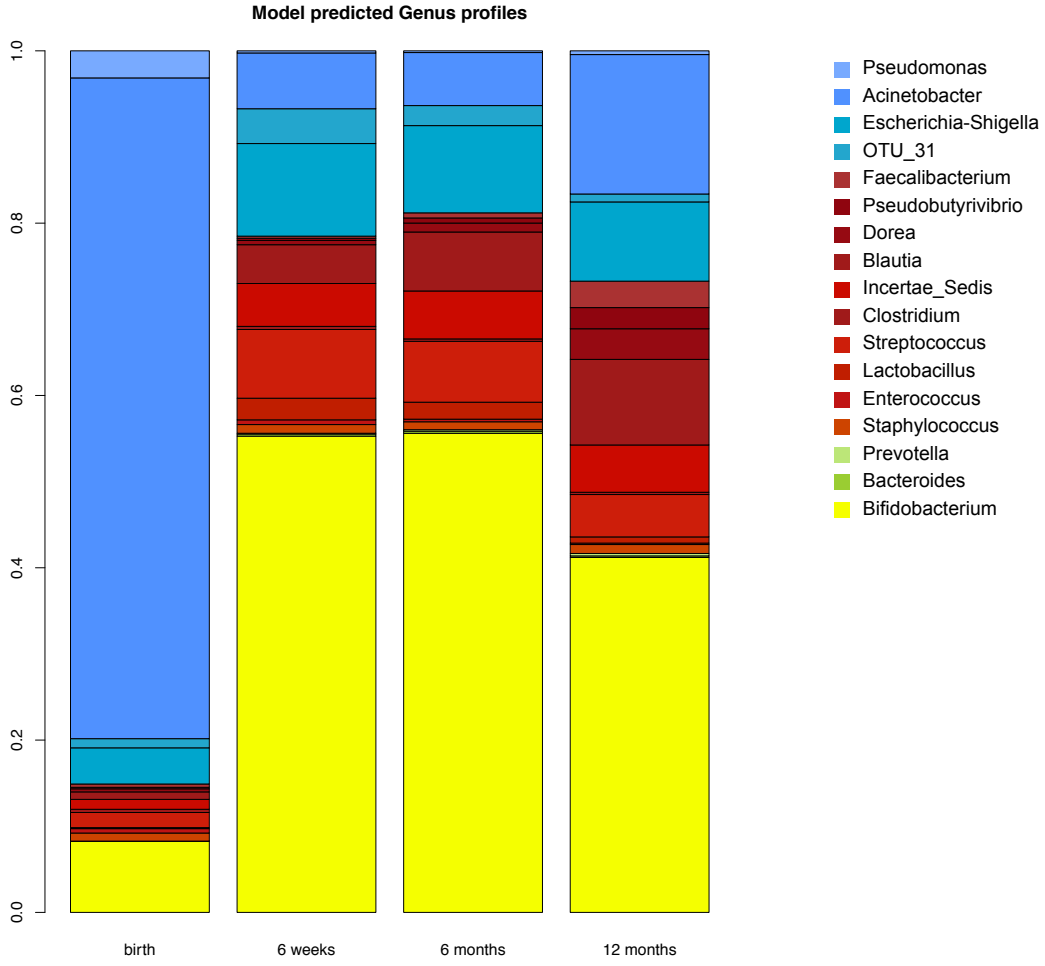


Figure 3.14. Genus level bacterial profiles of the infant stool samples from birth to 12 months.
The relative abundances of the bacterial genera in the infant stool samples at different time points were visualized by bar plots. Each bar represents all the subjects at that time-point, and each coloured box a bacterial taxon.

3.4.4. External factors contributing towards the composition of the infant gut bacterial profile

3.4.4.1. The effect of breastfeeding on the infant stool bacterial profile

Breastfed infants were defined as infants who did not receive formula at any of the time points, while formula-fed infants were defined as infants who received formula only. In this study, five main bacterial taxa were shown to be affected by breastfeeding, with only two of them being significant. The phylum *Actinobacteria* was found in significantly higher amounts for infants that were breastfed in the first year of life (p= 8.569442e-05) (Figure 3.15 A). The phylum *Proteobacteria* was also found in significantly lower amounts for breastfed infants (p=

3.480703e-02) (Figure 3.15 E). The phylum *Bacteroidetes* was found in higher amounts at the age of 12 months in infants that were not breastfed, though this was not significant (Figure 3.15 B). The genus *Bifidobacterium* was found in larger amounts for breastfed infants, while *Firmicutes* were in lower proportions (Figure 3.20 C, D).

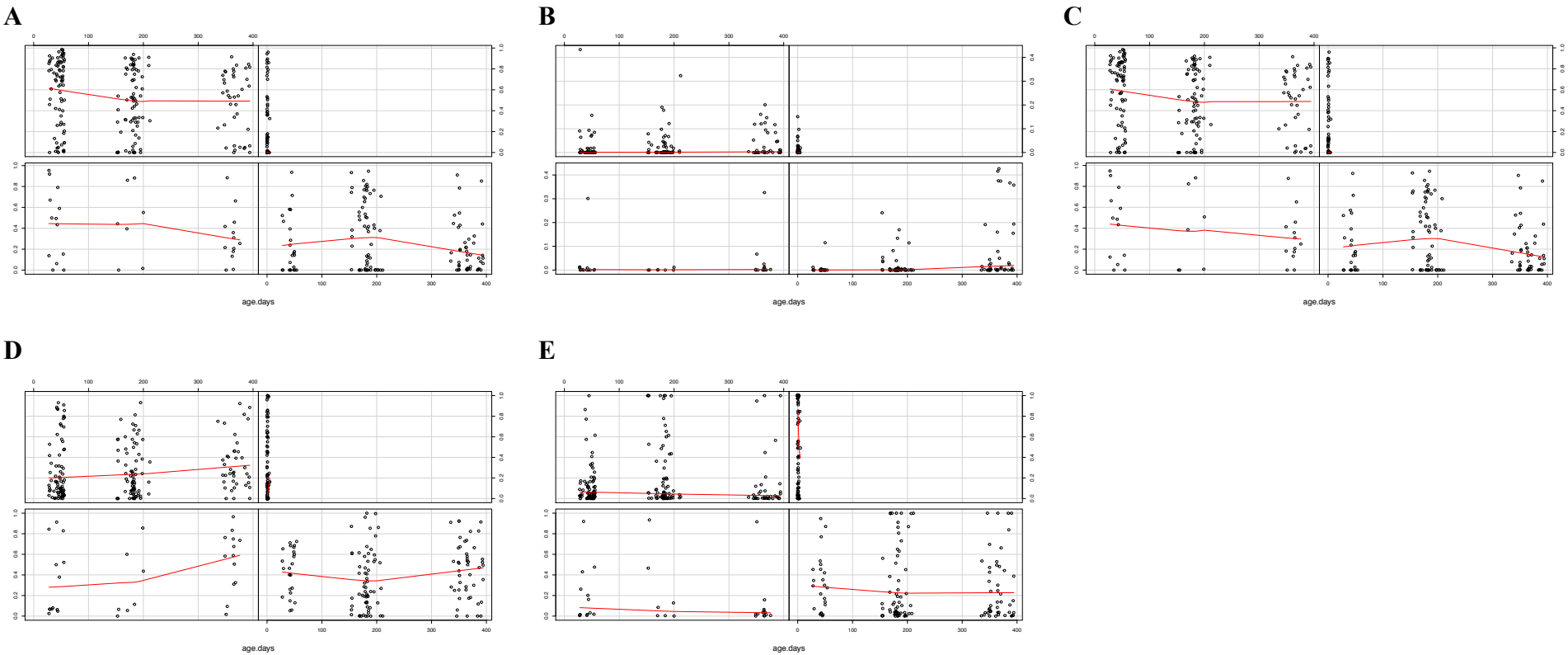


Figure 3.15. Bacterial Taxa affected by breastfeeding. (A) *Actinobacteria* **(B)** *Bacteroidetes* **(C)** *Bifidobacterium* **(D)** *Firmicutes* **(E)** *Proteobacteria*. The x-axis shows the age in days and the y-axis shows the relative abundance of the bacteria. The bottom left – infants with no breastfeeding data recorded, bottom right –formula fed, top left – breastfed, top right – taxa at birth for breastfed infants. The red line is a scatterplot smoother indicating the trend in that subset.

3.4.4.2. The effect of the use of electricity on the infant stool bacterial profile

Only three taxa were affected by electricity as the primary cooking source in this study. At birth, *Proteobacteria* was significantly lower in infants from homes that used electricity as energy sources for cooking (p= 3.525485e-04) (additional data which not shown). *Actinobacteria* was significantly increased (p= 8.130155e-09), in infants where electricity was the main cooking source, especially at six weeks of age (Figure 3.16 A). *Bifidobacterium* was increased in infants, especially at six weeks of age where electricity was the main cooking source, though this was not significant (Figure 3.21 B). At 12 months of age, the proportion of *Actinobacteria*, *Bifidobacterium* and *Proteobacteria* in infants whose households used electricity was similar to those that did not.

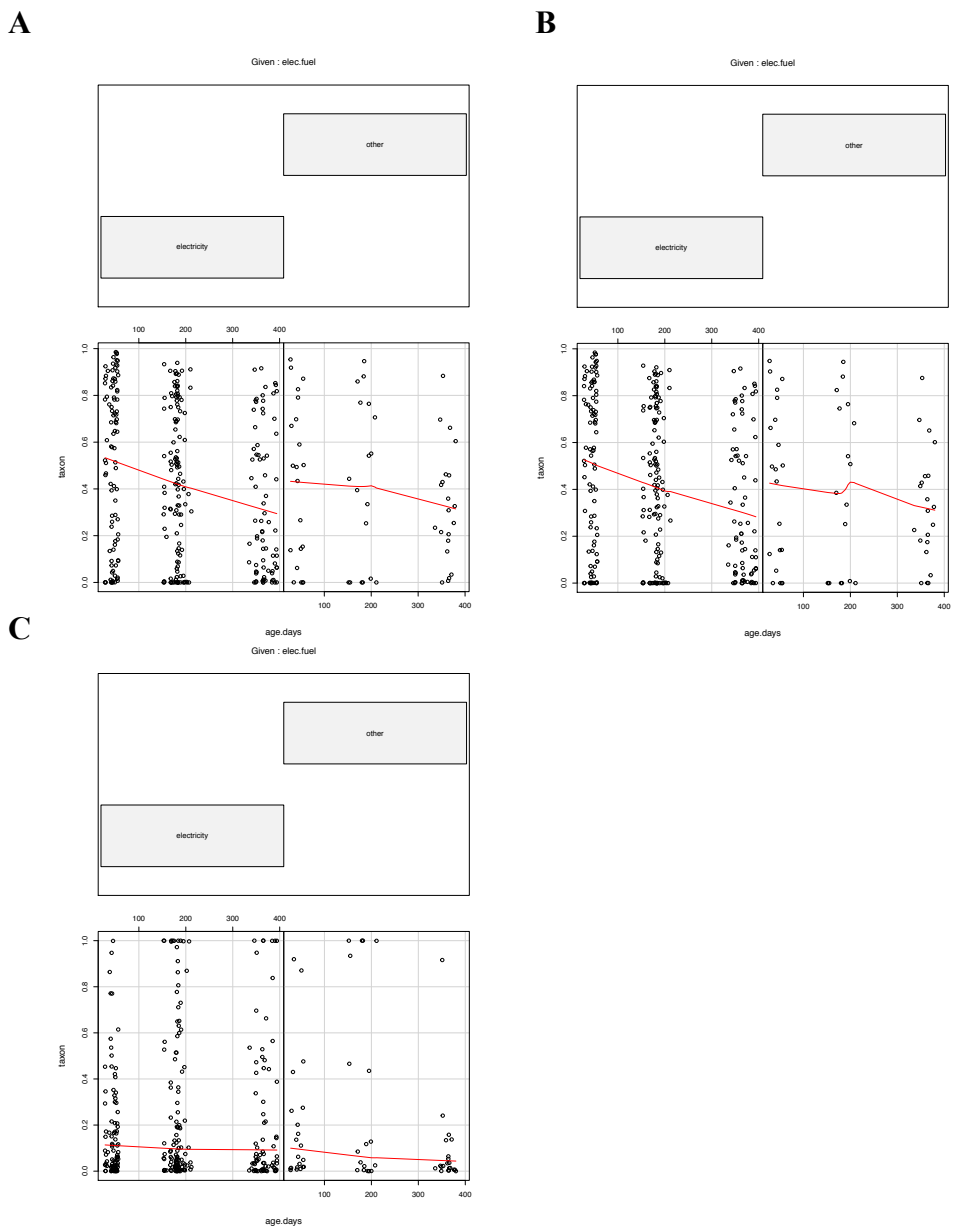


Figure 3.16. Bacterial Taxa affected by main cooking fuel (electricity) in the household. (A) *Actinobacteria* **(B)** *Bifidobacterium* **(C)** *Proteobacteria.* The x-axis shows the age in days and the y-axis shows the relative abundance of the bacteria. The left – electricity, right – other fuels. The red line is a scatterplot smoother indicating the trend in that subset.

3.4.4.3. Additional factors affecting the infant stool bacterial profile

Infants not exposed to cats had significantly higher proportions of *Proteobacteria* primarily between the ages six and 12 months (p= 0.01229328) (Figure 3.17 A). At six and 12 months, infants not exposed to dogs had significantly lower *Proteobacteria* in their stool (p= 0.001300965) (Figure 3.17 B). Infants that stayed in homes with four to eight household members had significantly higher amounts of *Proteobacteria* in their stool throughout the first

year of life (p= 0.02571465) (Figure 3.17 C). Household members were defined as other people who stayed in the house besides the mother, these included siblings and other adults. However, at 12 months of age infants with older siblings had decreased amounts of *Proteobacteria* in their stool, though this decrease was not significant (Figure 3.17 D). The effect of having other household members as well as older siblings was also analysed separately, since siblings are usually in closer contact with infants, and are likely to influence their bacterial profile in early years. The increase in the amount of *Firmicutes* was significantly pronounced in rural areas (p= 0.01690662), in comparison to urban areas and townships (Figure 3.17 E).

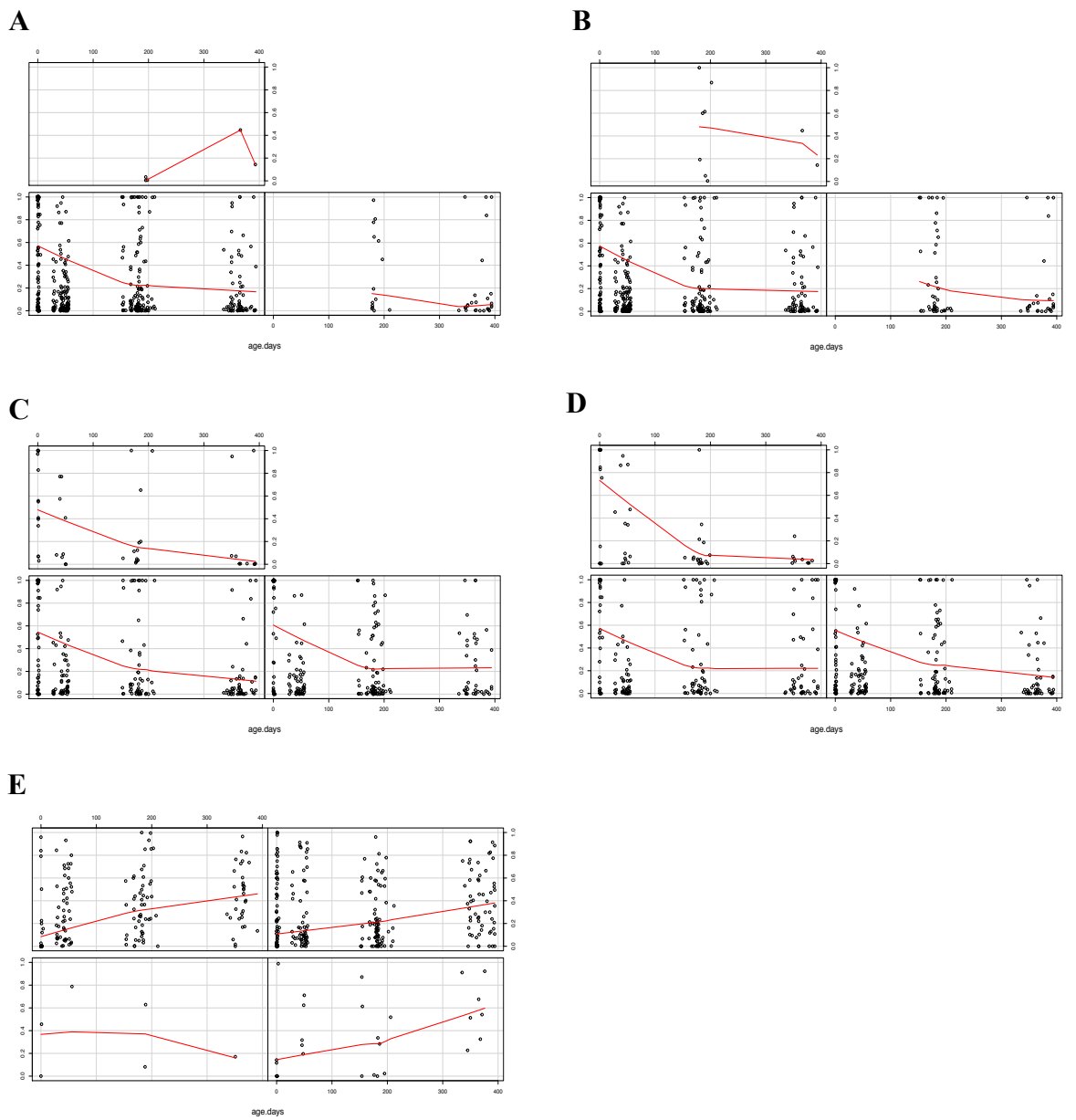


Figure 3.17. Risk factors that affect the proportions of bacterial members of the infant stool bacteria. The x-axis shows the age in days and the y-axis shows the relative abundance of the bacteria. **(A)** Cats and *Proteobacteria*. The bottom left – exposed to 0 cats, bottom right – exposed to 0-2 cats, top left – exposed to 2-6 cats. **(B)** Dogs and *Proteobacteria.* The bottom left – exposed to 0 dogs, bottom right – exposed to 0-2 dogs, top left – exposed to 2-6 dogs. **(C)** Household members and *Proteobacteria.* The bottom left – exposed to no household members, bottom right – exposed to 4-8 household members, top left – exposed to 8-20 household members. **(D)** Total number of children mother has and *Proteobacteria.* The bottom left – exposed to 0 siblings, bottom right – exposed to 0-2 siblings, top left – exposed to 2-5 siblings. **(E)** Living area and *Firmicutes.* The bottom left – living area type unknown, bottom right – rural living area, top left – township living area, top right – urban living area. The red line is a scatterplot smoother indicating the trend in that subset.

3.4.4.4. Factors affecting the infant stool diversity

Three main risk factors were found to significantly affect the diversity of the infant stool bacteria. Infants with older mothers had higher diversity in their stool (p= 0.0073) (Figure 3.18

A). Diversity in the infant stool samples also increased with age, but this progression was slower in babies that were shorter at birth (birth length < 40cm), (p= 0.00018) (Figure 3.18 B). For infants aged six months and under, diversity decreased as birth weight increased (p= 0.032). However, at 12 months the opposite is true (Figure 3.18 C).

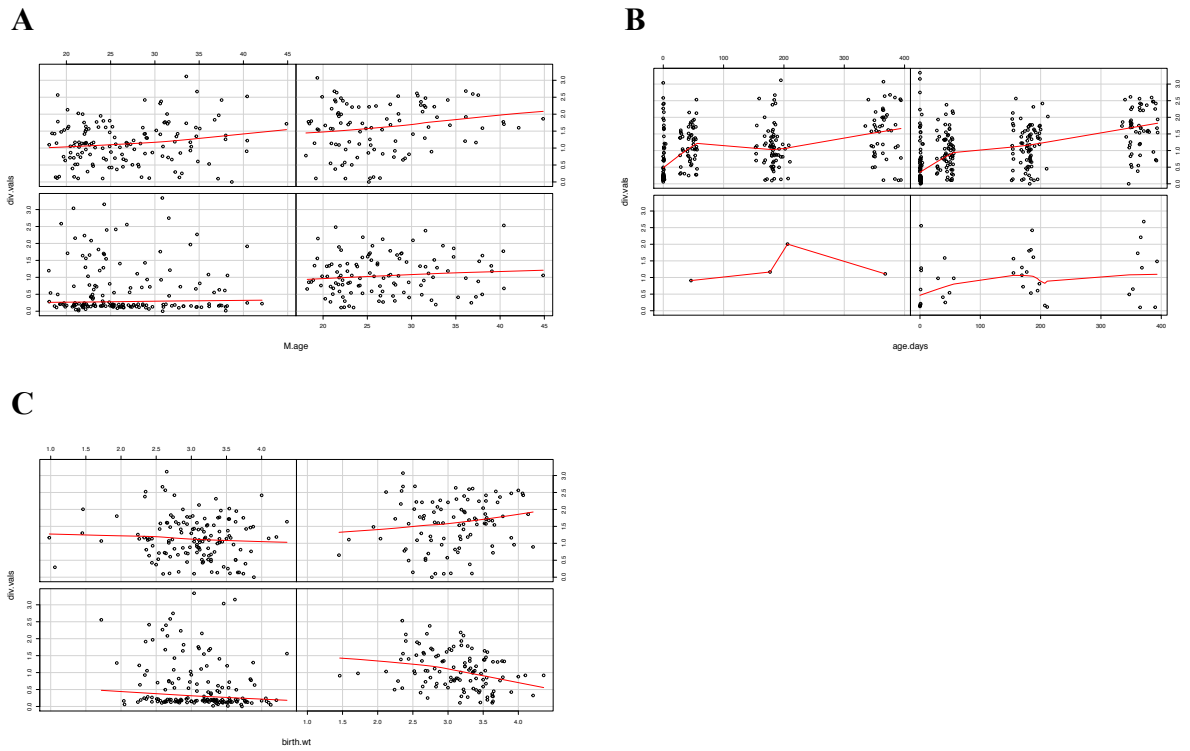


Figure 3.18. Risk factors affecting the diversity of the infant stool bacteria.
(A) Maternal age. The x-axis shows the maternal age and the y-axis shows the diversity. The bottom left – birth (infant), bottom right – 6 weeks (infant), top left – 6 months (infant), top right – 12 months (infant). **(B)** Birth length. The x-axis shows the age in days and the y-axis shows the diversity. The bottom left – 30-40cm, bottom right – 40-45cm, top left – 45-50cm, top right – 50-60cm. **(C)** Birth weight. The x-axis shows the age in days and the y-axis shows the diversity. The bottom left - birth, bottom right – 6 weeks, top left – 6 months, top right – 12 months. The red line is a scatterplot smoother indicating the trend in that subset.

3.4.5. Infant wheezing in the first year of life

3.4.5.1. The effect of wheezing on the composition of infant stool bacteria

Log ratio biplots of the infant stool samples (both aspirate and passed stools), based on wheezing status from birth to 12 months did not appear to form any distinct clusters both at phylum and genus level (Figures 3.19 and 3.20). In the first year of life, the infants stool bacterial profiles at phylum level did not cluster according to the different wheezing outcomes

(Figure 3.19). At genus level from birth to one year, the infant stool bacterial profiles were also randomly dispersed in relation to the wheezing outcome (Figure 3.20).

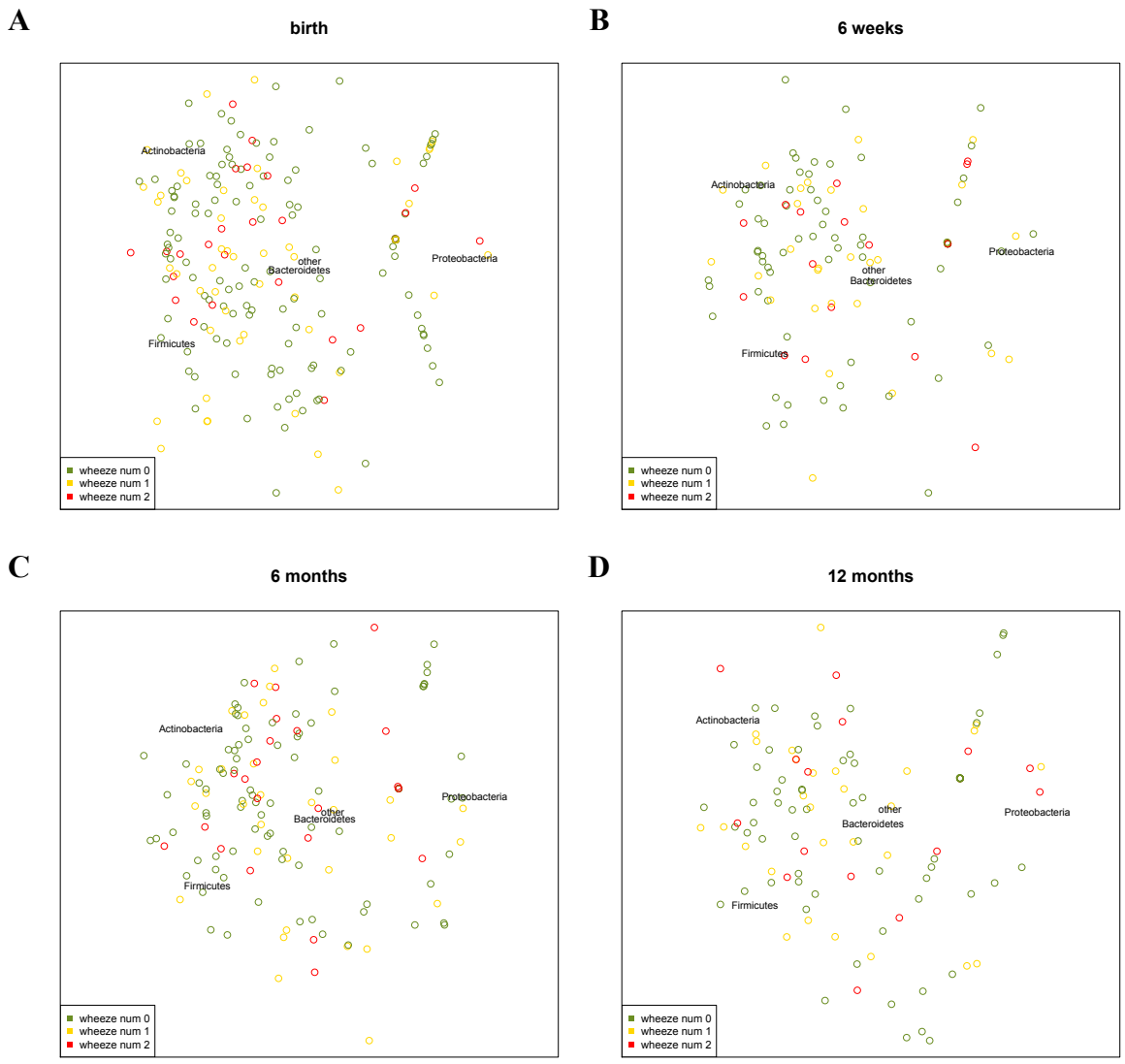


Figure 3.19. Log ratio biplots of infant stool samples at phylum level in relation to wheezing sampled at birth, six weeks, six months and 12 months.
Stool samples of infants that did not wheeze are shown in green, those that wheezed once in gold and those that wheezed twice, or more are shown in red.

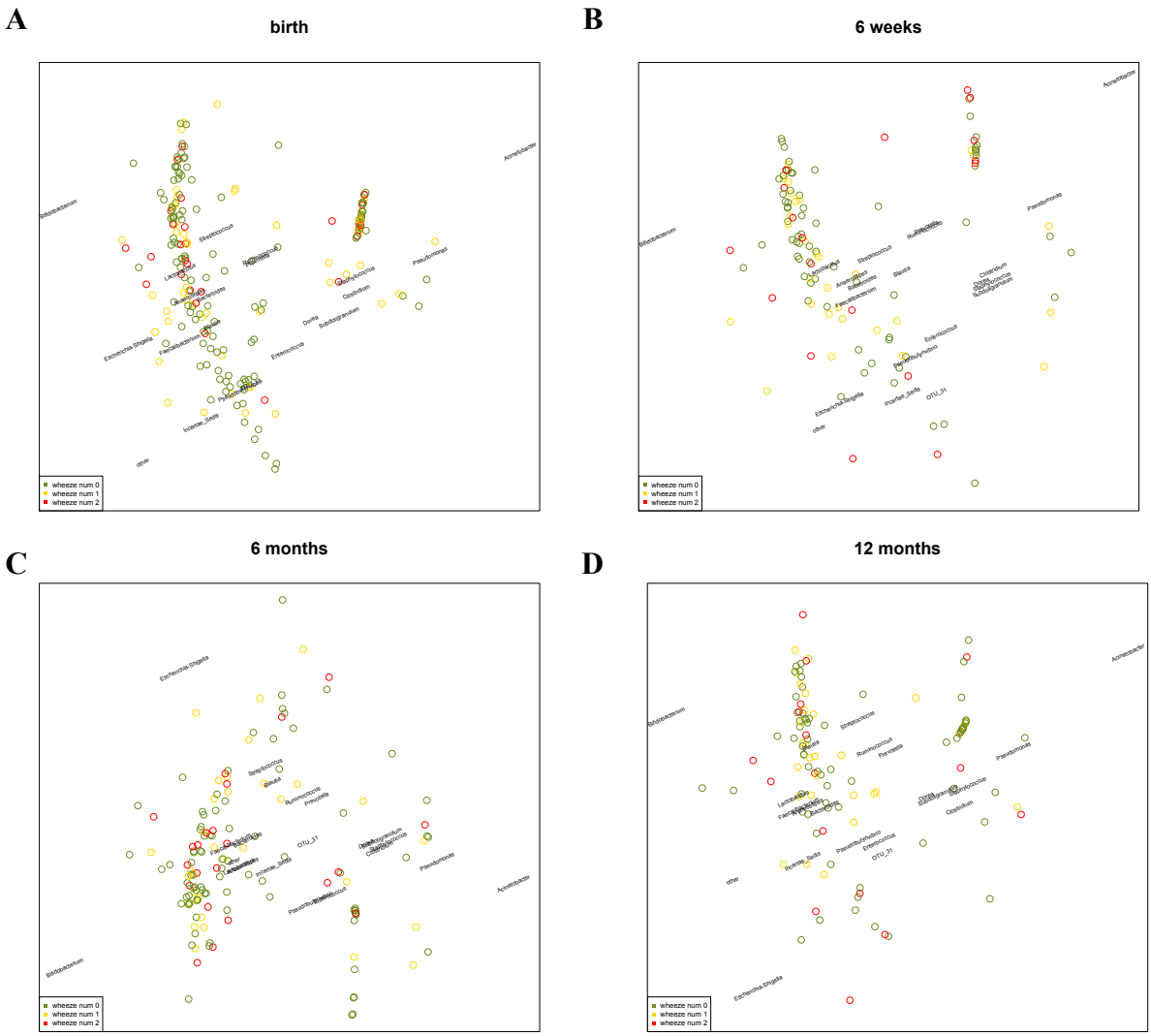


Figure 3.20. Log ratio biplots of the infant stool samples at genus level in relation to wheezing sampled at birth, six weeks, six months and 12 months.
Stool samples of infants that did not wheeze are shown in green, those that wheezed once in gold and those that wheezed twice or more are shown in red.

3.4.5.2. The association between wheezing and infant bacteria

The proportions of the *Lactobacillales* in infants who wheezed recurrently during the first year of life was more than those who wheezed once during the same period. Infants who wheezed once, with an onset of about seven days, had no change in the proportion of *Lactobacillales* from birth to 12 months of age (Figure 3.21). On the other hand, infants who wheezed recurrently, with the same onset of seven days had increased proportions of *Lactobacillales* from birth to 12 months (Figure 3.21). Infants with a wheezing onset of both 140 and about

210 days, did not have differences in proportions of *Lactobacillales* between those that wheezed once and those that wheezed recurrently (Figure 3.21). Infants who wheezed two or more times in the first year of life had significantly higher proportions of *Lactobacillales* present in their stool (p=0.04083). The proportion of *Proteobacteria* decreased from birth to 12 months in infants that did not wheeze (Figure 3.22). The amount of *Proteobacteria* increased at 12 months if the infants had wheezed before (Figure 3.22).

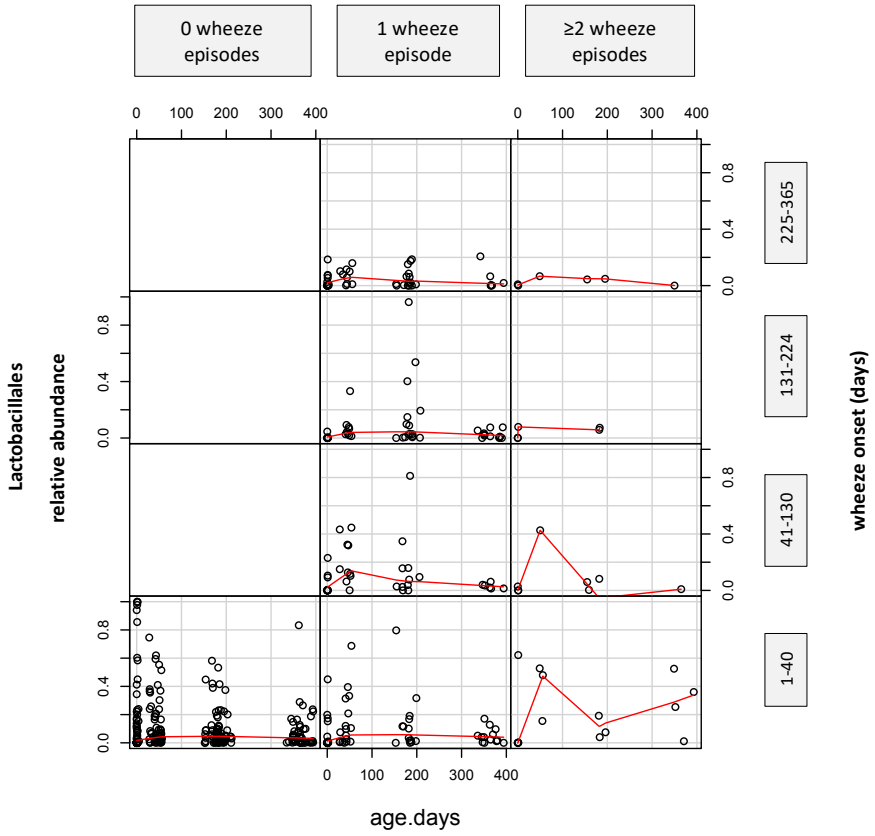


Figure 3.21. Coplot of bacteria at order level and wheezing variables (*Lactobacillales*).
The x-axis represents the age of infants in days. The y-axis represents the relative abundance of the *Lactobacillales* present in the infant samples. The three vertical panels represent the different wheezing outcomes; 0 wheeze, 1 wheeze, ≥2 wheeze episodes. The four horizontal panels represent the four wheezing onset groups; zero to 40 days, 41 to 130 days, 131 to 224 days, 225 to 365 days. The red line is a scatterplot smoother indicating the trend in that subset.

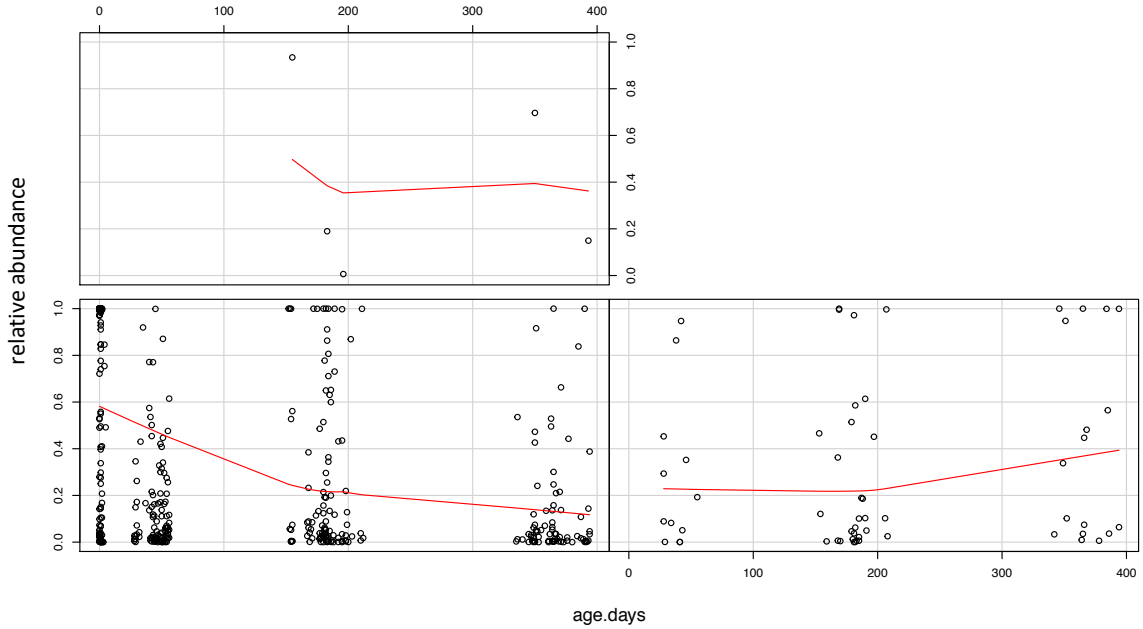


Figure 3.22. Coplot of bacteria at phylum level and wheezing variables (*Proteobacteria).*
The x-axis represents the age of infants in days. The y-axis represents the relative abundance of the *Proteobacteria* present in the infant samples. The bottom left – 0 wheezing episodes, bottom right – wheezing episode, top left – 2-4 wheezing episodes. The red line is a scatterplot smoother indicating the trend in that subset.

3.5. Discussion

The principal aim of the study was to investigate the association between the stool bacteria and the development of recurrent wheezing in infants. In this study, the composition of certain stool bacteria was associated with the development of wheezing illness. The log-ratio biplots of the overall bacterial profiles, generated for the wheezing outcome in this study did not show any distinct clusters. This means that infants with different wheezing outcomes did not have significant differences within their stool bacterial profiles, to group separately. Thus, the overall infant bacterial profiles were fairly similar between the two wheezing outcomes; infants that wheezed once and infants that wheezed two or more times in the first year of life. However, upon generation of the coplots, one bacterial taxon (*Lactobacillales)* was seemingly significantly associated with the development of wheezing though there were no clear differences in children with late wheeze onset. A study by Amberbir and colleagues reported

that the presence of *Lactobacillus* was not associated with increased risk of allergic symptoms such as wheezing, which is contrary to the results found in our study (Amberbir *et al.*, 2011). A study by Penders and colleagues also reported that colonization by *Lactobacillus* was not associated with atopic manifestations, including wheezing (Penders *et al.*, 2007). Penders and colleagues also concluded that development of recurrent wheeze was associated with colonization by *C. difficile* in early infancy (Penders *et al.*, 2007). A more recent study by Salas and colleagues observed that at one year of age wheezing infants had over-representations of the *Lachnospiraceae*, *Rikenellaceae* and *Ruminococcaceae* families (Salas *et al.*, 2014). This is different to what was found in the Canadian studies, as they reported the reduction in proportion of the FLVR bacteria in wheezing infants at three months of age, and much less apparent differences at one year (Arrieta *et al.*, 2015). Salas and colleagues also reported that infants with two or more wheezing episodes had under-represented amounts of the *Clostridium* genus (Salas *et al.*, 2014). More recent studies have investigated the combined contribution of both the bacterial and the fungal components of the stool in the development of wheezing and asthma later on in life (Arrieta *et al.*, 2017; Fujimura *et al.*, 2017). Another study by Arrieta and colleagues carried out in rural Ecuador identified different bacterial and fungal groups responsible for the development of wheezing in infants (Arrieta *et al.*, 2017). Ecuadorian infants with increased proportions of *Streptococcus* and *Bacteroides* species and reduced *Bifidobacterium* and *Ruminococcus gnavus* developed wheezing later on in infancy (Arrieta *et al.*, 2017). The wheezing infants from the Ecuador study also had reduced *Saccharomycetales* and increased *Pichia kudriavzevii* in their three month stools (Arrieta *et al.*, 2017). Arrieta and colleagues were able to demonstrate the importance of the fungal component in the development of infant wheezing in this Ecuadorian cohort (Arrieta *et al.*, 2017). The vast differences between the results of the two studies by Arrieta and colleagues in Canada and Ecuador bring forth the idea that different bacterial and fungal species are responsible for the

development of wheezing in different regions of the world. In spite of this information obtained, we suspect that the observed differences between the proportions of *Lactobacillales* in infants who wheezed and those who did not were due to outliers and few data points for recurrent wheezers. We assume that more infants that have wheezed recurrently would need to be included to increase the statistical strength of the results for this particular objective.

To our knowledge, no infant bacterial studies have been carried out using both aspirated and passed stool. Patel and colleagues are one of the groups that used gastric aspirates as the sample of choice for 16S rRNA-based Denaturing Gradient Gel Electrophoresis, but did not use passed stool samples or measure the difference if any (Patel *et al.*, 2015). For this study, mean beta diversities were measured using Whittaker (β_w) and Simpson's β_{sim} formulas, and similar patterns were observed for the aspirated and passed stool samples (Simpson, 1943; Whittaker, 1960). Aspirated stool samples were observed to be similar to passed stool samples, because of the clustering of the two sample types in the multidimensional scaling plots. Statistically significant differences between aspirated and passed stool samples were due to concentration differences. The purity of the passed stool samples in this study was significantly higher than that of the aspirated stool samples. The lower purity values in the aspirate samples indicate the presence of contaminants like phenols and proteins that would skew this value (Thermo Scientific). In longitudinal studies, passed stool can sometimes be difficult to obtain at specific time points, especially in infants. Given that aspirated stool can be collected at any time, this makes this sample type favourable for stool bacterial studies going forward. This study was able to confirm that the use of aspirate stool samples is sufficient for use in 16S rRNA sequencing analysis. Therefore, aspirated stool samples can be used in gastrointestinal microbiome studies with confidence that the results will be representative of passed stool.

In this study, the bacterial profile at six months was dominated by the phylum *Actinobacteria,* followed by *Firmicutes, Proteobacteria* and *Bacteroidetes*. A study by Thompson and colleagues showed a similar profile, except that their infants had a higher proportion of *Bacteroidetes* than *Proteobacteria* (Thompson *et al.*, 2015). Another study reported a six month infant bacterial profile that was dominated by *Firmicutes*, followed by *Actinobacteria* and *Bacteroidetes* (González *et al.*, 2013). A study by Azad and colleagues, which had infants that were sampled at four months of age supported the observations made in this study, though the proportions were slightly different (Azad *et al.*, 2013). This study as well as the study by Azad and colleagues had a similar objective; the effect of risk factors such as presence of siblings and pets on the infant stool composition (Azad *et al.*, 2013).

The overall diversity of the infant stool bacteria in this study increased over time, as is reported by other researchers (Konya *et al.*, 2014; Arrieta *et al.*, 2015; Azad *et al.*, 2015, 2016; Bäckhed *et al.*, 2015). This increase in diversity was expected, because as infants grow they are exposed to more sources of bacteria (Frese and Mills, 2015). In this study, the infant stool diversity was positively affected by formula feeding. Previous studies addressing feeding methods in infants, have reported conflicting results. Some studies have shown formula fed infants to have higher diversity (Thompson *et al.*, 2015), while other studies have shown breastfed infants to have higher diversity (Gewolb, Schwalbe and Taciak, 1999). The data gathered from this study, as well as other studies confirms that different feeding methods cause increase in the infant stool diversity, though the overall profiles might differ.

Diet or feeding method has been shown to influence the infant stool bacterial profile during the first year of life (Palmer *et al.*, 2007; Madan *et al.*, 2012; Yatsunenko *et al.*, 2012; Groer *et al.*, 2014). Breastfed infants in this cohort had decreased amounts of *Firmicutes* and

Proteobacteria and increased proportions of *Actinobacteria,* specifically *Bifidobacterium.* This profile is similar to that of the breast milk, which is lower in *Proteobacteria* and higher in the phyla *Actinobacteria*, suggesting that certain bacterial taxa are acquired through this feeding method (Penders *et al.*, 2006; Palmer *et al.*, 2007; Turroni *et al.*, 2012; Di Mauro *et al.*, 2013; Torrazza and Josef, 2013). The infants that were formula fed were found to have increased proportions of the phylum *Bacteroidetes* (Penders *et al.*, 2006; Adlerberth and Wold, 2009). The differences observed between breastfed and formula fed infants supports the notion that the feeding method significantly affects or alters the infant stool composition.

Crowding entails any of the following: attendance at a daycare centre, presence of siblings or a large number of household members. In this study, infants in homes with more than five household members had higher amounts of *Proteobacteria* in their stool throughout the first year of life. A study focusing on the respiratory microbiome supports the increase in the phylum *Proteobacteria* in infants exposed to crowding via daycare centres and presence of older siblings (Bosch *et al.*, 2017). Similarly, in this study, infants with siblings also exhibited higher amounts of *Proteobacteria* compared to those without. The presence of older siblings has been associated with a reduced risk of developing atopic diseases, including asthma and food allergy (Strachan, 1989; Koplin *et al.*, 2012). Previous studies investigating the association between having older siblings and the infant gut bacterial profile have mostly observed increased amounts of the phyla *Firmicutes* and *Actinobacteria.* This profile was represented by the increased proportions of *Bifidobacteria* and *Staphylococcus aureus*, in conjunction with delayed colonization of the taxon *Clostridium* (Penders *et al.*, 2006; Adlerberth *et al.*, 2007; Lindberg *et al.*, 2011). In this study, the presence of siblings was not associated with increased amounts of the phyla *Firmicutes* and *Actinobacteria* as found in other studies.

Exposure of infants to pets has been associated with a reduced risk of developing atopic diseases, including asthma and food allergy (Strachan, 1989; Hesselmar *et al.*, 1999; Koplin *et al.*, 2012). Infants in this study exposed to cats had significantly lower proportions of the phylum *Proteobacteria* from six to 12 months when compared to unexposed infants. Whereas infants in this study exposed to dogs had higher proportions of the phylum *Proteobacteria*. Nermes and colleagues observed differences in the proportions of *Bifidobacterium* species in pet exposed and non-exposed infants (Nermes *et al.*, 2013). Tun and colleagues reported increased richness of the phylum *Firmicutes* in the stool of pet exposed infants (Tun *et al.*, 2017). From the results of this current study and others by different research groups it can be noted that pet exposure affects different bacterial groups. This study showed a significant increase in the phylum *Firmicutes* in infants living in rural areas. Similarly, other studies have observed gut bacteria dominated by *Firmicutes* in individuals from rural areas (Tyakht *et al.*, 2013; Martinez *et al.*, 2015). In contrast, De Filippo and colleagues found a depletion of *Firmicutes* in children from rural areas in Burkina Faso (De Filippo *et al.*, 2010).

3.6. Conclusion

Aspirate and passed stool samples were found to be very similar and indistinguishable. There were no differences in the overall stool bacterial profiles of the infants with the different wheezing outcomes. Recurrent wheezing was significantly associated with the composition of *Lactobacillales*, though this association was due to outliers and few data points. Some of the limitations of this study included inadequate number of recurrent wheezers, incomplete clinical data due to poor collection and incomplete sample collection for infants. We envisage that a more extensive sample size is required to be able to obtain statistically significant results.

3.7. References

Adlerberth, I. *et al.* (2007) 'Gut microbiota and development of atopic eczema in 3 European birth cohorts', *Journal of Allergy and Clinical Immunology*, 120(2), pp. 343–350. doi: 10.1016/j.jaci.2007.05.018.

Adlerberth, I. and Wold, A. E. (2009) 'Establishment of the gut microbiota in Western infants', *Acta Paediatrica, International Journal of Paediatrics*, 98(2), pp. 229–238. doi: 10.1111/j.1651-2227.2008.01060.x.

Akay, H. K. *et al.* (2014) 'The relationship between bifidobacteria and allergic asthma and/or allergic dermatitis: A prospective study of 0-3 years-old children in Turkey', *Anaerobe*, 28, pp. 98–103. doi: 10.1016/j.anaerobe.2014.05.006.

Amberbir, a. *et al.* (2011) 'Effects of Helicobacter pylori, geohelminth infection and selected commensal bacteria on the risk of allergic disease and sensitization in 3-year-old Ethiopian children', *Clinical and Experimental Allergy*, 41(10), pp. 1422–1430. doi: 10.1111/j.1365-2222.2011.03831.x.

Andrews, S. (2010) *FastQC: A quality control tool for high throughput sequence data*.

Arrieta, M. *et al.* (2015) 'Early infancy microbial and metabolic alterations affect risk of childhood asthma', *Science Translational Medicine*, 7(307), p. 307ra152-307ra152. doi: 10.1126/scitranslmed.aab2271.

Arrieta, M. C. *et al.* (2017) 'Associations between infant fungal and bacterial dysbiosis and childhood atopic wheeze in a nonindustrialized setting', *Journal of Allergy and Clinical Immunology*. doi: 10.1016/j.jaci.2017.08.041.

Azad, M. B. *et al.* (2013) 'Gut microbiota of healthy Canadian infants: Profiles by mode of delivery and infant diet at 4 months', *CMAJ*, 185(5). doi: 10.1503/cmaj.121189.

Azad, M. B. *et al.* (2015) 'Infant gut microbiota and food sensitization: associations in the first year of life', *Clinical & Experimental Allergy*, 45(3), pp. 632–643. doi: 10.1111/cea.12487.

Azad, M. B. *et al.* (2016) 'Impact of maternal intrapartum antibiotics, method of birth and breastfeeding on gut microbiota during the first year of life: A prospective cohort study', *BJOG: An International Journal of Obstetrics and Gynaecology*, 123(6), pp. 983–993. doi: 10.1111/1471-0528.13601.

Bäckhed, F. *et al.* (2015) 'Dynamics and stabilization of the human gut microbiome during the first year of life', *Cell Host and Microbe*, 17(5), pp. 690–703. doi: 10.1016/j.chom.2015.04.004.

Bisgaard, H. *et al.* (2011) 'Reduced diversity of the intestinal microbiota during infancy is associated with increased risk of allergic disease at school age', *Journal of Allergy and Clinical Immunology*, 128. doi: 10.1016/j.jaci.2011.04.060.

Blanken, M. O. *et al.* (2013) 'Respiratory syncytial virus and recurrent wheeze in healthy preterm infants.', *The New England journal of medicine*, 368(19), pp. 1791–9. doi: 10.1056/NEJMoa1211917.

Bogaert, D. *et al.* (2011) 'Variability and diversity of nasopharyngeal microbiota in children: A metagenomic analysis', *PLoS ONE*, 6(2). doi: 10.1371/journal.pone.0017035.

Bosch, A. A. T. M. *et al.* (2017) 'Maturation of the infant respiratory microbiota,

environmental drivers, and health consequences. A prospective cohort study.', *American journal of respiratory and critical care medicine*, 196(12), pp. 1582–1590. doi: 10.1164/rccm.201703-0554OC.

Caporaso, J. G. *et al.* (2010) 'QIIME allows analysis of high-throughput community sequencing data', 7(5), pp. 335–336. doi: 10.1038/nmeth.f.303.QIIME.

Caporaso, J. G. *et al.* (2011) 'Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample.', *Proceedings of the National Academy of Sciences of the United States of America*, 108 Suppl(Supplement_1), pp. 4516–22. doi: 10.1073/pnas.1000080107.

Claassen-Weitz, S. *et al.* (2018) 'HIV-exposure, early life feeding practices and delivery mode impacts on faecal bacterial profiles in a South African birth cohort', *Scientific reports*, 8(1), p. 5078. doi: 10.1038/s41598-018-22244-6.

Edgar, R. C. (2010) 'Search and clustering orders of magnitude faster than BLAST', *Bioinformatics*, 26(19), pp. 2460–2461. doi: 10.1093/bioinformatics/btq461.

Edgar, R. C. (2013) 'UPARSE: highly accurate OTU sequences from microbial amplicon reads.', *Nature methods*, 10(10), pp. 996–998. doi: 10.1038/nmeth.2604.

De Filippo, C. *et al.* (2010) 'Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa.', *Proceedings of the National Academy of Sciences of the United States of America*, 107(33), pp. 14691–6. doi: 10.1073/pnas.1005963107.

Fox, J. and Weisberg, S. (2011) 'An {R} companion to applied regression'. Available at: http://socserv.socsci.mcmaster.ca/jfox/Books/Companion.

Frese, S. A. and Mills, D. A. (2015) 'Birth of the infant gut microbiome: Moms deliver twice!', *Cell Host and Microbe*, 17(5), pp. 543–544. doi: 10.1016/j.chom.2015.04.014.

Fujimura, K. E. *et al.* (2017) 'Neonatal gut microbiota associates with childhood multi–sensitized atopy and T–cell differentiation', *Nature medicine*, 22(10), pp. 1187–1191. doi: 10.1038/nm.4176.Neonatal.

Gewolb, I. H., Schwalbe, R. S. and Taciak, V. L. (1999) 'Stool microflora in extremely low birthweight infants', *Arch Dis Child Fetal Neonatal Ed*, 80, pp. F167–F173.

González, R. *et al.* (2013) 'Breast milk and gut microbiota in african mothers and infants from an area of high hiv prevalence.', *PLoS ONE*, 8(11), pp. 1–9. doi: 10.1371/journal.pone.0080299.

Gosalbes, M. J. *et al.* (2013) 'Meconium microbiota types dominated by lactic acid or enteric bacteria are differentially associated with maternal eczema and respiratory problems in infants', *Clinical & Experimental Allergy*, 43(2), pp. 198–211. doi: 10.1111/cea.12063.

Greenacre, M. J. (2010) *Biplots in practice*, *Fundacion BBVA*.

Groer, M. W. *et al.* (2014) 'Development of the preterm infant gut microbiome: A research priority', *Microbiome*, 2(1), pp. 1–8. doi: 10.1186/2049-2618-2-38.

Hastie, T. (2013) 'Fits a principal curve in arbitrary dimension'.

Hesselmar, B. *et al.* (1999) 'Does early exposure to cat or dog protect against later allergy development?', *Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology*, 29(5), pp. 611–617.

Hughes, J. B. *et al.* (2016) 'Counting the uncountable : statistical approaches to estimating microbial diversity mini review counting the uncountable : statistical approaches to

estimating microbial diversity', *Applied and environmental microbiology*, 10(1), pp. 4399–4406. doi: 10.1128/AEM.67.10.4399.

Hurlbert, S. H. (1971) 'The nonconcept of species diversity: a critique and alternative parameters.', *Ecology*, 52(4), pp. 577–586. doi: 10.2307/1934145.

Kalliomäki, M. *et al.* (2001) 'Distinct patterns of neonatal gut microflora in infants in whom atopy was and was not developing', *Journal of Allergy and Clinical Immunology*, 107, pp. 129–134. doi: 10.1067/mai.2001.111237.

Konya, T. *et al.* (2014) 'Associations between bacterial communities of house dust and infant gut', *Environmental Research*. Elsevier, pp. 25–30. doi: 10.1016/j.envres.2014.02.005.

Koplin, J. J. *et al.* (2012) 'Environmental and demographic risk factors for egg allergy in a population-based study of infants.', *Allergy*, 67(11), pp. 1415–1422. doi: 10.1111/all.12015.

Kusel, M. M. H. *et al.* (2007) 'Early-life respiratory viral infections, atopic sensitization, and risk of subsequent development of persistent asthma', *Journal of Allergy and Clinical Immunology*, 119, pp. 1105–1110. doi: 10.1016/j.jaci.2006.12.669.

de Leeuw, J. and Mair, P. (2009) 'Multidimensional Scaling Using Majorization: SMACOF in R', *Journal of Statistical Software*, 31(3), pp. 1–30. Available at: tp://www.jstatsoft.org/v31/i03/.

Lennon, J.J., Koleff, P., Greenwood, J.J.D., Gaston, K. J. (2001) 'The geographical structure of British bird distributions: diversity, spatial turnover and scale', *Journal of Animal Ecology*, 70, pp. 966–979.

Liaw, A. and Wiener, M. (2018) 'Classification and regression by random forest', *R News*, 2(3), pp. 18–22. Available at: https://cran.r-project.org/doc/Rnews/.

Lindberg, E. *et al.* (2011) 'Effect of lifestyle factors on Staphylococcus aureus gut colonization in Swedish and Italian infants', *Clinical Microbiology and Infection*. European Society of Clinical Microbiology and Infectious Diseases, 17(8), pp. 1209–1215. doi: 10.1111/j.1469-0691.2010.03426.x.

Ly, N. P. *et al.* (2006) 'Recurrent wheeze in early childhood and asthma among children at risk for atopy.', *Pediatrics*, 117(6), pp. e1132–e1138. doi: 10.1542/peds.

Madan, J. C. *et al.* (2012) 'Nomal neonatal microbiome variation in relation to environmental factors, infection, and allergy', *Current Opinion in Pediatrics*, 24(6), pp. 753–759. doi: 10.1097/MOP.0b013e32835a1ac8.Normal.

Maechler, M. *et al.* (2015) 'Cluster: Cluster analysis basics and extensions'.

Martinez, F. D. *et al.* (1995) 'Asthma and wheezing in the first six years of life', 332(3).

Martinez, I. *et al.* (2015) 'The gut microbiota of rural papua new guineans: composition, diversity patterns, and ecological processes.', *Cell reports*, 11(4), pp. 527–538. doi: 10.1016/j.celrep.2015.03.049.

Di Mauro, A. *et al.* (2013) 'Gastrointestinal function development and microbiota', *Italian Journal of Pediatrics*, 39(1), p. 1. doi: 10.1186/1824-7288-39-15.

McMurdie, P. J. and Paulson, J. N. (2016) 'biomformat:An interface package for the BIOM file format', *https://github.com/joey711/biomformat/, http://biom-format.org/.*

Morgan, X. C. and Huttenhower, C. (2012) 'Chapter 12: Human Microbiome Analysis', *PLoS Computational Biology*, 8(12). doi: 10.1371/journal.pcbi.1002808.

Murray, C. S. *et al.* (2005) 'Fecal microbiota in sensitized wheezy and non-sensitized non-wheezy children: a nested case–control study.', *Clinical & Experimental Allergy*, 35(6), pp. 741–745. Available at: 10.1111/j.1365-2222.2005.02259.x.

Nermes, M. *et al.* (2013) 'Perinatal pet exposure, faecal microbiota, and wheezy bronchitis: is there a connection?', *ISRN Allergy*, 2013, p. 827934. doi: 10.1155/2013/827934.

Nriagu, J. *et al.* (1999) 'Prevalence of asthma and respiratory symptoms in south-central Durban, South Africa.', *European journal of epidemiology*, 15(8), pp. 747–755. doi: 10.1023/A:1007653709188.

Oksanen, J., Blanchet, F. and Kindt, R, et al. (2013) 'Vegan: Community Ecology Package'.

Palmer, C. *et al.* (2007) 'Development of the human infant intestinal microbiota', *PLoS Biology*, 5(7), pp. 1556–1573. doi: 10.1371/journal.pbio.0050177.

Panettieri, R. A. *et al.* (2008) 'Natural history of asthma: Persistence versus progression-does the beginning predict the end?', *Journal of Allergy and Clinical Immunology*, 121(3), pp. 607–613. doi: 10.1016/j.jaci.2008.01.006.

Patel, K. *et al.* (2015) 'Trends and determinants of gastric bacterial colonization of preterm neonates in a NICU setting', *PLoS ONE*, 10(7), pp. 1–21. doi: 10.1371/journal.pone.0114664.

Penders, J. *et al.* (2006) 'Factors influencing the composition of the intestinal microbiota in early infancy.', *Pediatrics*, 118(2), pp. 511–521. doi: 10.1542/peds.2005-2824.

Penders, J. *et al.* (2007) 'Gut microbiota composition and development of atopic manifestations in infancy: the KOALA Birth Cohort Study.', *Gut*, 56(5), pp. 661–667. doi: 10.1136/gut.2006.100164.

Pylro, V. S. *et al.* (2014) 'Data analysis for 16S microbial profiling from different benchtop sequencing platforms', *Journal of Microbiological Methods*, 107, pp. 30–37. doi: 10.1016/j.mimet.2014.08.018.

Remes, S. T. *et al.* (2001) 'Dog exposure in infancy decreases the subsequent risk of frequent wheeze but not of atopy', *Journal of Allergy and Clinical Immunology*, 108, pp. 509–515. doi: 10.1067/mai.2001.117797.

Revelle, W. (2018) 'psych: Procedures for Psychological, Psychometric, and Personality Research'. Available at: https://cran.r-project.org/package=psych.

Salas, S. M. *et al.* (2014) 'Infant gut microbiota and the development of wheeze in early childhood', *Allergy, Asthma & Clinical Immunology*, 10(Suppl 1), p. A35. doi: 10.1186/1710-1492-10-S1-A35.

Shannon, C. E. (1948) 'A mathematical theory of communication', *The Bell System Technical Journal*, 27(July 1928), pp. 379–423. doi: 10.1145/584091.584093.

Simpson, G. G. (1943) 'Mammals and the nature of continents', *American Journal of Science*, 241, pp. 1–31.

Smuts, H. E., Workman, L. J. and Zar, H. J. (2011) 'Human rhinovirus infection in young African children with acute wheezing.', *BMC infectious diseases*, 11(1), p. 65. doi: 10.1186/1471-2334-11-65.

Stiemsma, L. T. *et al.* (2016) 'Shifts in Lachnospira and Clostridium sp. in the 3-month stool microbiome are associated with preschool age asthma', *Clinical Science*, 130(23), pp. 2199–2207. doi: 10.1042/CS20160349.

Strachan, D. P. (1989) ‘Hay fever, hygiene, and household size.’, *BMJ (Clinical research ed.)*, 299, pp. 1259–1260. doi: 10.1136/bmj.299.6710.1259.

Sunyer, J. *et al.* (2001) ‘Prenatal risk factors of wheezing at the age of four years in Tanzania.’, *Thorax*, 56(4), pp. 290–295. doi: 10.1136/thorax.56.4.290.

Taussig, L. M. *et al.* (2003) ‘Tucson children’s respiratory study: 1980 to present’, *Journal of Allergy and Clinical Immunology*, 111(4), pp. 661–675. doi: 10.1067/mai.2003.162.

Thompson, A. L. *et al.* (2015) ‘Milk- and solid-feeding practices and daycare attendance are associated with differences in bacterial diversity, predominant communities, and metabolic and immune function of the infant gut microbiome.’, *Frontiers in cellular and infection microbiology*, 5(February), p. 3. doi: 10.3389/fcimb.2015.00003.

Torrazza, R. M. and Josef, N. (2013) ‘The altered gut microbiome and necrotizing Enterocolitis’, *Clinics in Perinatology*, 40(1), pp. 93–108. doi: 10.1016/j.clp.2012.12.009.The.

Tun, H. M. *et al.* (2017) ‘Exposure to household furry pets influences the gut microbiota of infants at 3–4 months following various birth scenarios’, *Microbiome*, 5(1), p. 40. doi: 10.1186/s40168-017-0254-x.

Turroni, F. *et al.* (2012) ‘Diversity of bifidobacteria within the infant gut microbiota’, *PLoS ONE*, 7(5), pp. 20–24. doi: 10.1371/journal.pone.0036957.

Tyakht, A. V. *et al.* (2013) ‘Human gut microbiota community structures in urban and rural populations in Russia’, *Nature Communications*, 4, pp. 1–9. doi: 10.1038/ncomms3469.

Vael, C. *et al.* (2008) ‘Early intestinal Bacteroides fragilis colonisation and development of asthma’, *BMC Pulmonary Medicine*, 8, p. 19. doi: 10.1186/1471-2466-8-19.

Vael, C. *et al.* (2011) ‘Denaturing gradient gel electrophoresis of neonatal intestinal microbiota in relation to the development of asthma.’, *BMC Microbiology*, 11(1), p. 68. doi: 10.1186/1471-2180-11-68.

Venables, W. N. and Ripley, B. D. (2002) ‘Modern applied statistics with S (fourth.)’, *Retrieved from http://www.stats.ox.ac.uk/pub/MASS4*.

Wang, Q. *et al.* (2007) ‘Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy.’, *Applied and environmental microbiology*, 73(16), pp. 5261–5267. doi: 10.1128/AEM.00062-07.

Whittaker, R. H. (1960) ‘Vegetation of the Siskiyou mountains, Oregon and California’, *Ecological Monographs*, 30, pp. 279–338.

Yatsunenko, T. *et al.* (2012) ‘Human gut microbiome viewed across age and geography’, *Nature*, 486(7402), pp. 222–227. doi: 10.1038/nature11053.

Zar, H. J. *et al.* (2015) ‘Investigating the early-life determinants of illness in Africa: the Drakenstein Child Health Study.’, *Thorax*, 70(6), pp. 592–594. doi: 10.1136/thoraxjnl-2014-206242.

3.8. Appendix

Appendix A

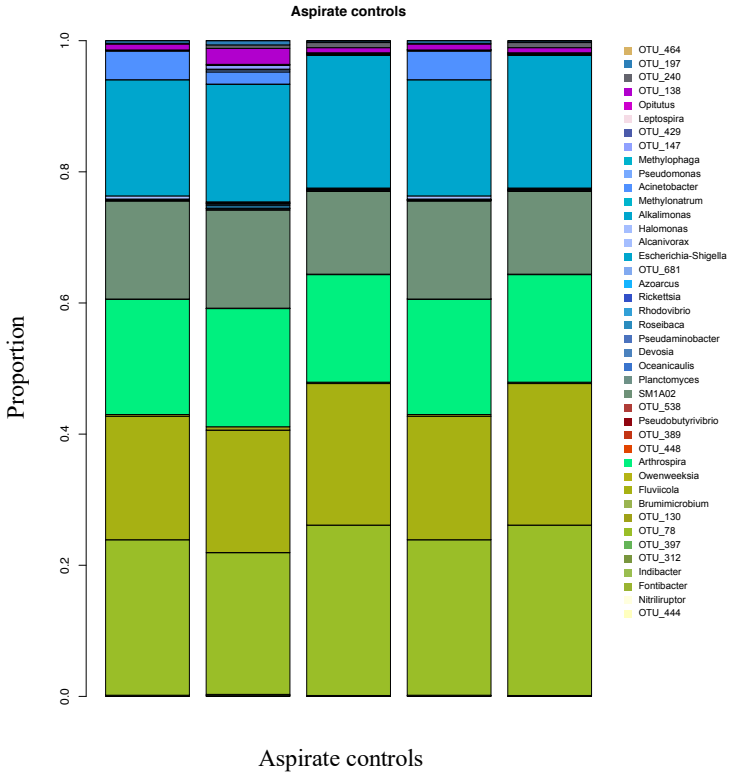


Figure S1. Genus level bacterial profiles of the of the aspirate controls in this study. The relative abundances of the bacterial genera in the aspirate controls were visualized by bar plots. Each bar on the x-axis represents an aspirate control, and each coloured box a bacterial taxon. The y-axis represents the proportions of the bacterial groups in the controls.

Appendix B
Infant Class Profiles over time

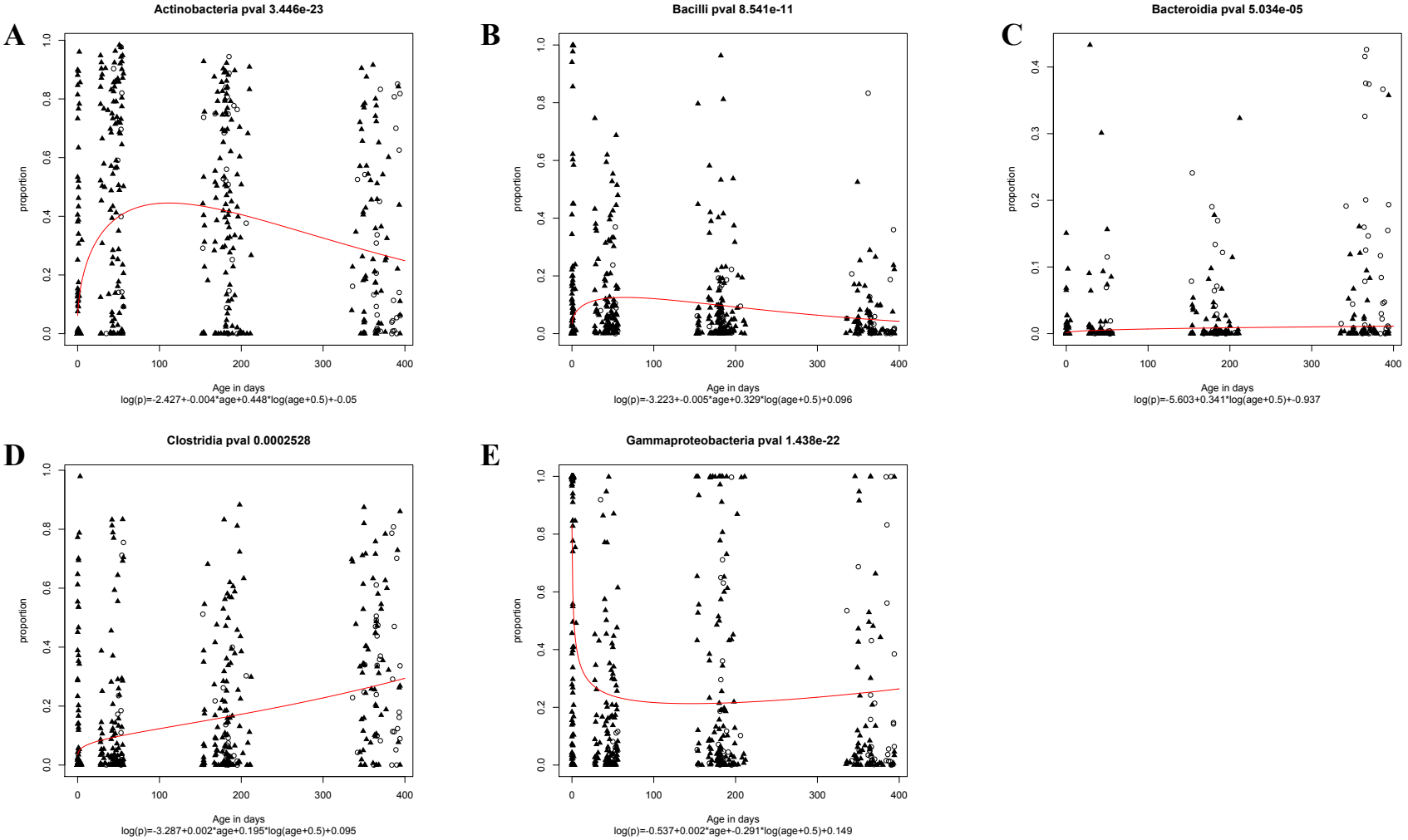


Figure S2. Relative abundance plots for the infants at class level from birth to 12 months. (A) *Actinobacteria* **(B)** *Bacilli* **(C)** *Bacteroidia* **(D)** *Clostridia* **(E)** *Gammaproteobacteria***.**

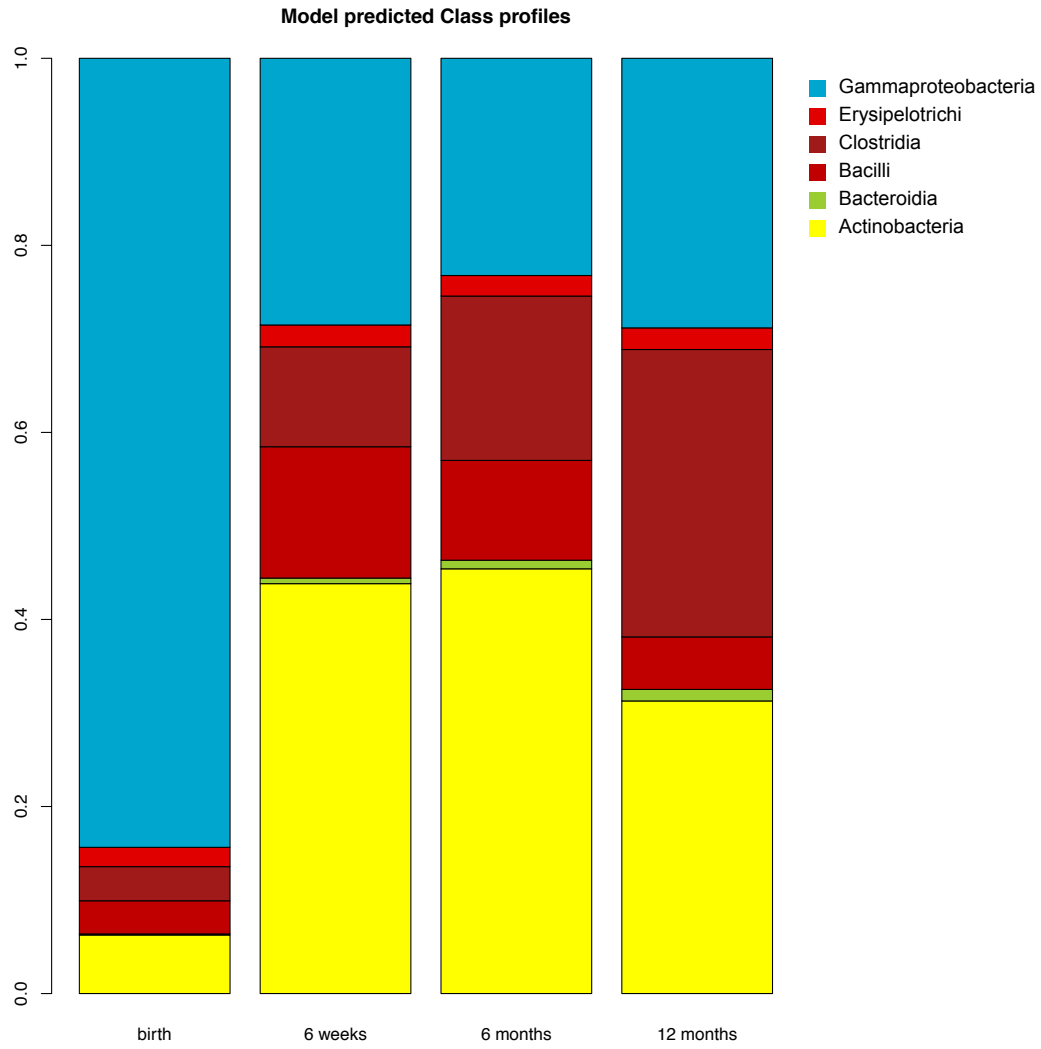


Figure S3. Class level bacterial profiles of the infant stool samples from birth to 12 months. The relative abundances of the bacterial phyla in the infant stool samples at different time points were visualized by bar plots. Each bar represents all the subjects at that time-point, and each coloured box a bacterial taxon.

Infant Order profiles over time

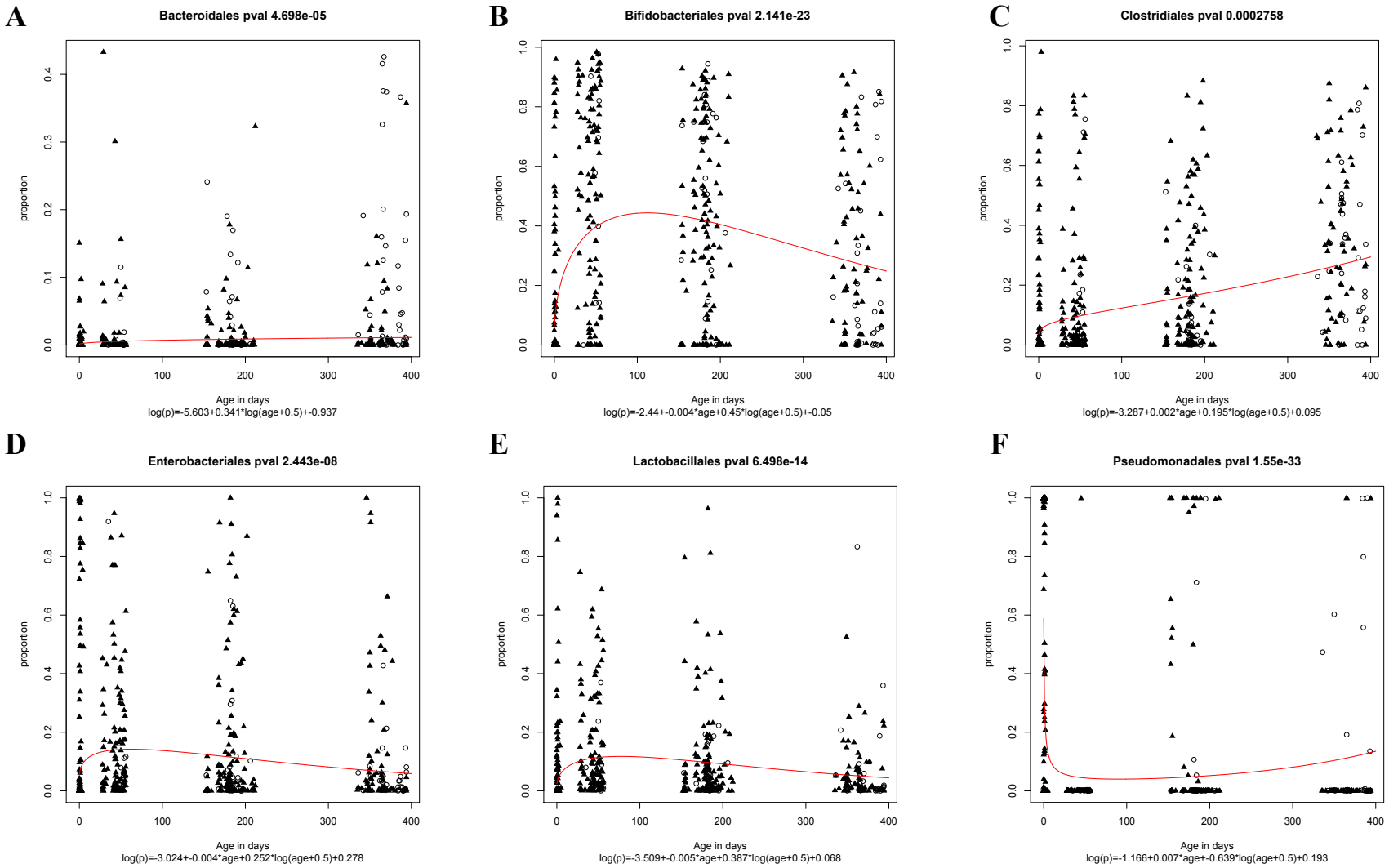


Figure S4. Relative abundance plots for the infants at order level from birth to 12 months. (A) *Bacteroidales* **(B)** *Bifidobacteriales* **(C)** *Clostridiales* **(D)** *Enterobacteriales* **(E)** *Lactobacillales* **(F)** *Pseudomonadales*.

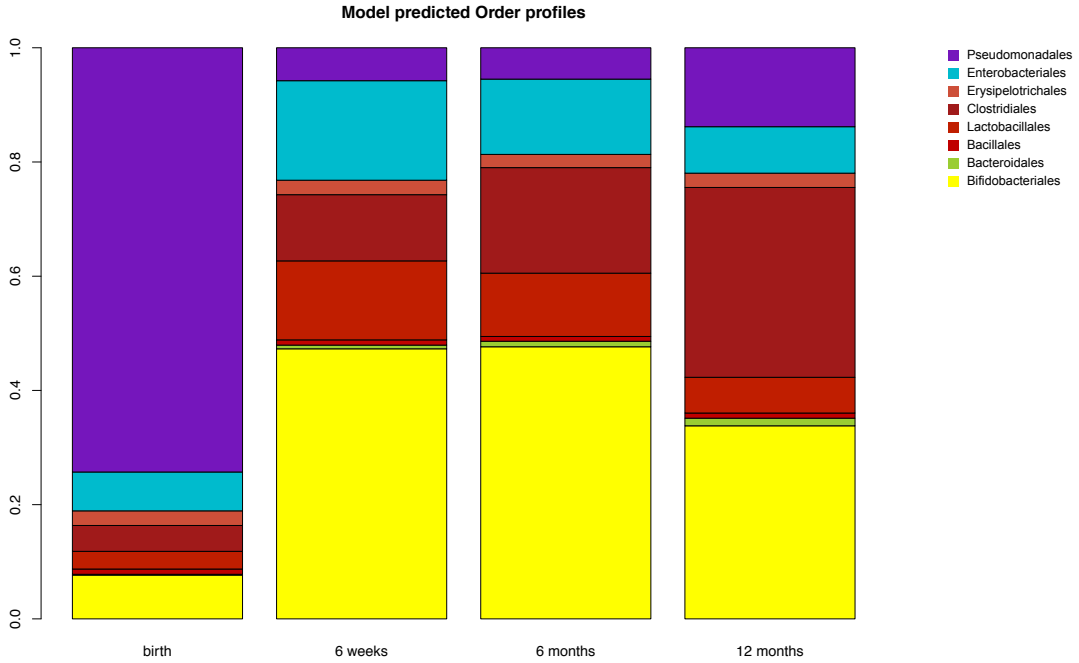
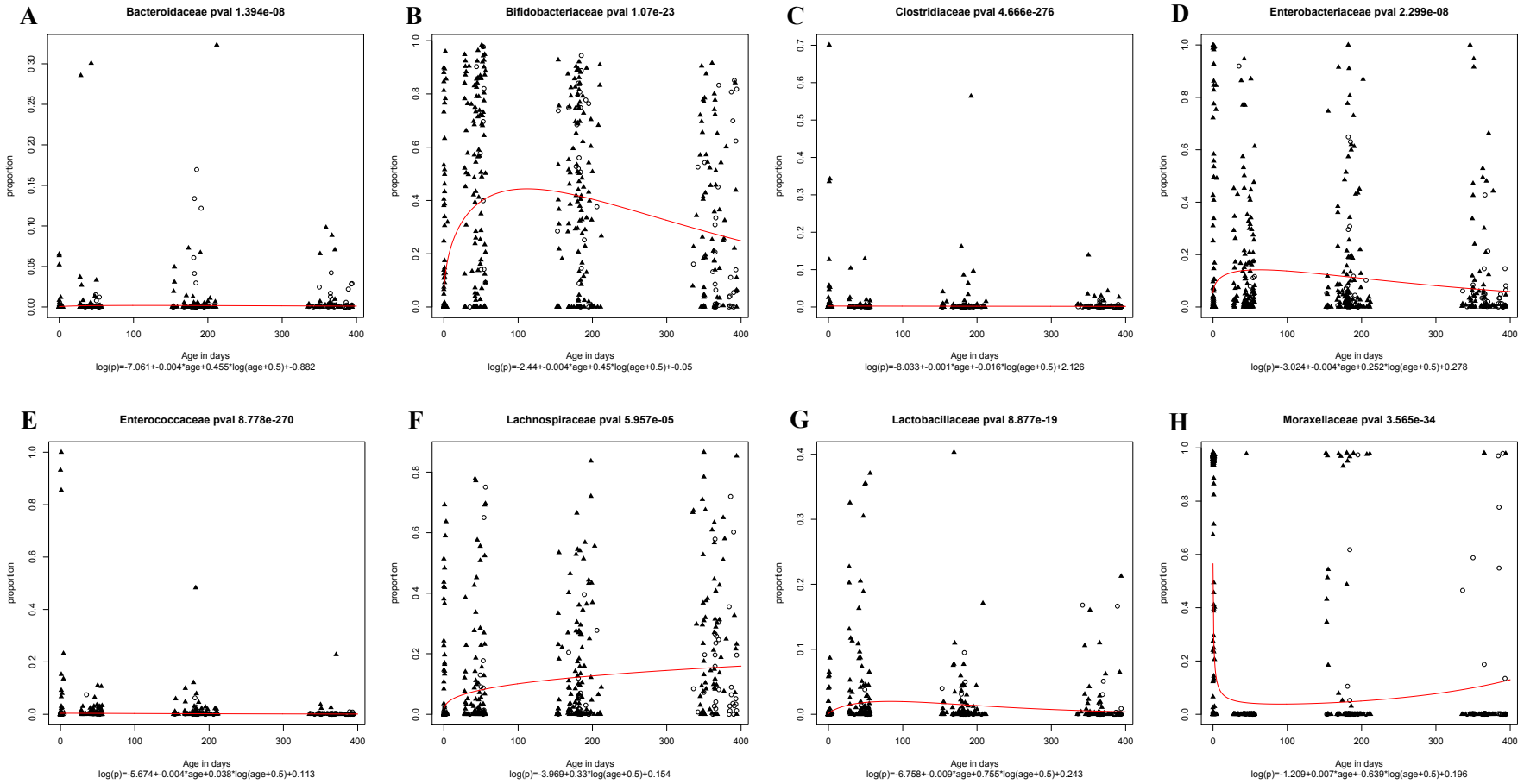


Figure S5. Order level bacterial profiles of the infant stool samples from birth to 12 months. The relative abundances of the bacterial orders in the infant stool samples at different time points were visualized by bar plots. Each bar represents all the subjects at that time-point, and each coloured box a bacterial taxon.

Infant Family profiles over time



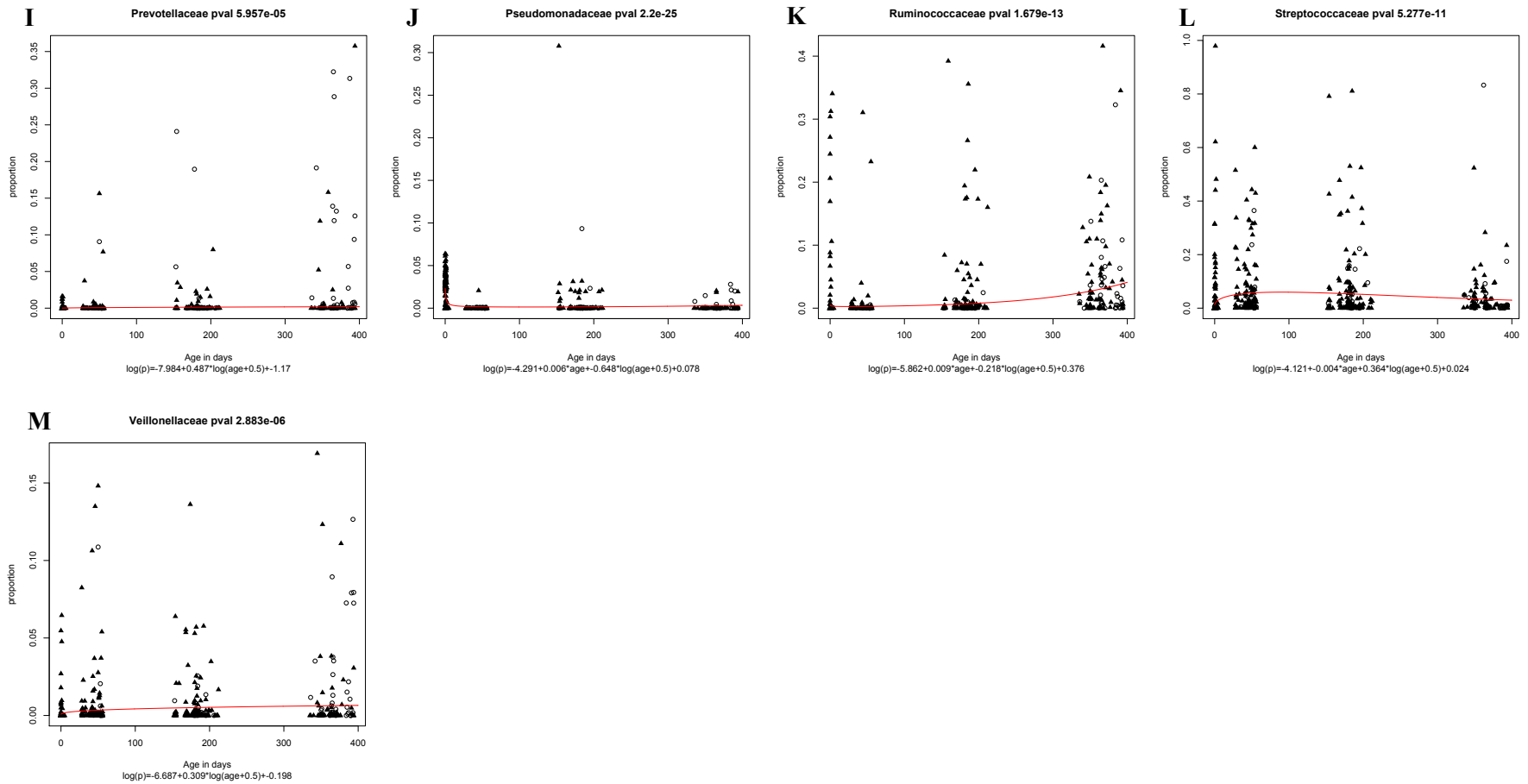


Figure S6. Relative abundance plots for the infants at phylum level from birth to 12 months. (A) *Bacteroidaceae* **(B)** *Bifidobacteriaceae* **(C)** *Clostridiaceae* **(D)** *Enterobacteriaceae* **(E)** *Enterococcaceae* **(F)** *Lachnospiraceae* **(G)** *Lactobacillaceae* **(H)** *Moraxellaceae* **(I)** *Prevotellaceae* **(J)** *Pseudomonadaceae* **(K)** *Ruminococcaceae* **(L)** *Streptococcaceae* **(M)** *Veillonellacea***.**

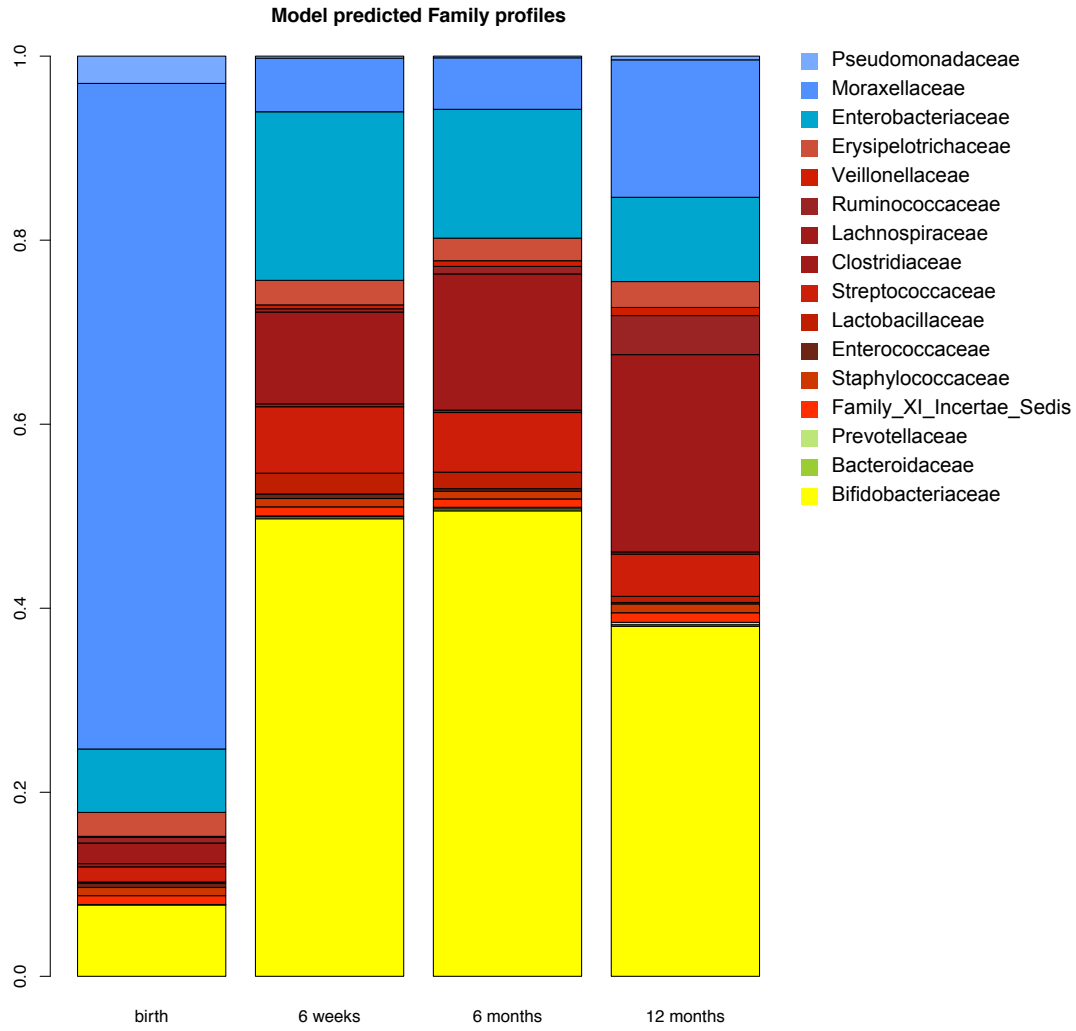
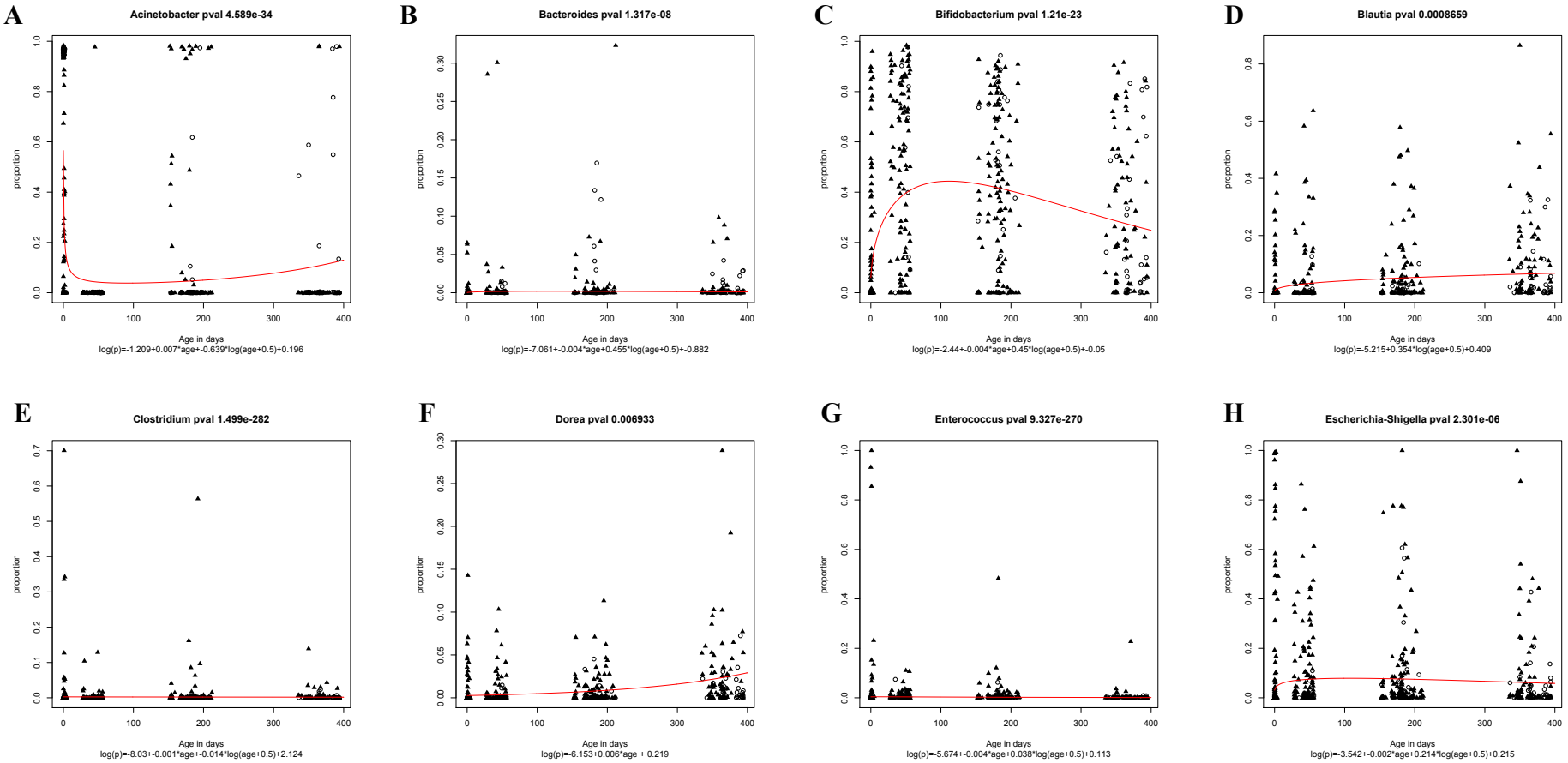


Figure S7. Family level bacterial profiles of the infant stool samples from birth to 12 months. The relative abundances of the bacterial families in the infant stool samples at different time points were visualized by bar plots. Each bar represents all the subjects at that time-point, and each coloured box a bacterial taxon.

Infant Genus profiles over time



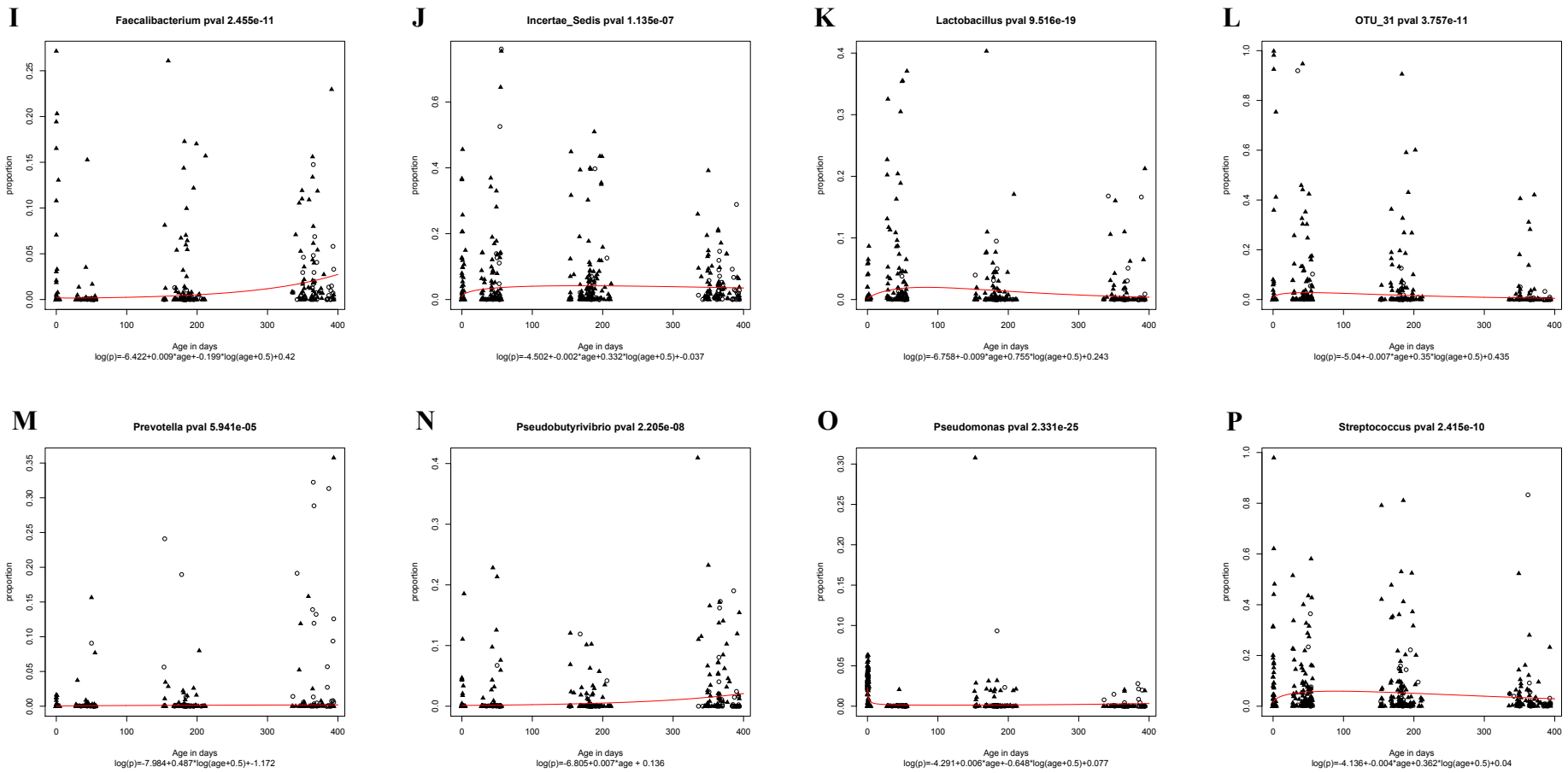


Figure S8. Relative abundance plots for the infants at genus level from birth to 12 months.
(A) *Acinetobacter* **(B)** *Bacteroides* **(C)** *Bifidobacterium* **(D)** *Blautia* **(E)** *Clostridium* **(F)** *Dorea* **(G)** *Enterococcus* **(H)** *Escherichia-Shigella* **(I)** *Faecalibacterium* **(J)** *Incertae Sedis* **(K)** *Lactobacillus* **(L)** *OTU 31* **(M)** *Prevotella* **(N)** *Pseudobutyrivibrio* **(O)** *Pseudomonas* **(P)** *Streptococcus*.
○: aspirated stool ▲ sample : passed stool samples

Appendix D
Ethics Approval Letter

UNIVERSITY OF CAPE TOWN
Faculty of Health Sciences
Human Research Ethics Committee

Room E52-24 Old Main Building
Groote Schuur Hospital
Observatory 7925
Telephone [021] 406 6338 • **Facsimile** [021] 406 6411
Email: sumayah.ariefdien@uct.ac.za
Website: www.health.uct.ac.za/fhs/research/humanethics/forms

12 October 2015

HREC REF: 748/2015

Dr M Kaba
Division of Medical Microbiology
Room 5.30 Level 5,
Falmouth Building-FHS

Dear Dr Kaba

PROJECT TITLE: THE STOOL MICROBIOTA AND INFANT WHEEZING ILLNESS – THE DRAKENSTEIN CHILD HEALTH STUDY, SOUTH AFRICA (Sub-study linked to 401/2009) MSc candidate – Ms M Ngwarai)

Thank you for submitting your study to the Faculty of Health Sciences Human Research Ethics Committee (HREC) for review.

It is a pleasure to inform you that the HREC has formally approved the above-mentioned study.

Approval is granted for one year until the 30th September 2016.

Please submit a progress form, using the standardised Annual Report Form if the study continues beyond the approval period. Please submit a Standard Closure form if the study is completed within the approval period.
(Forms can be found on our website: www.health.uct.ac.za/fhs/research/humanethics/forms)

We acknowledge that the following student: - Michelle Ngwarai is also involved in this project.

Please quote the HREC reference no in all your correspondence.

Please note that the ongoing ethical conduct of the study remains the responsibility of the principal investigator.

Yours sincerely

signature removed to avoid exposure online

PROFESSOR M BLOCKMAN
CHAIRPERSON, FHS HUMAN RESEARCH ETHICS COMMITTEE
Federal Wide Assurance Number: FWA00001637.
Institutional Review Board (IRB) number: IRB00001938

Hrec/ref:748/2015

Chapter 4 - General Discussion

4.1. Discussion

The aims of this Masters' dissertation were to use next generation sequencing technology to describe the major components of the infant gut bacterial profile in the first year of life, identify major risk factors that influenced this composition and investigate the association between the gut bacteria and the development of recurrent wheezing in infants. This study was carried within the Drakenstein Child Health Study, which is a prospective birth cohort from a low socio-economic community that is found in the Western Cape of South Africa.

To the best of our knowledge, this is the first study to use two different stool sampling techniques, namely passed stool and aspirated stool, in infants for 16S rRNA sequencing. Aspirated stool was considered as a stool sampling option for infants, due to the difficulty experienced when trying to collect the sample on a specific day. The results from the initial statistical analysis showed that aspirated and passed stool samples were indistinguishable, hence proving the high level of similarity. This demonstrated that aspirated samples were good enough as an alternative for passed stool samples. Other studies have investigated the use of rectal swabs as an alternative to passed stool, and have observed minimal profile differences between the two sampling techniques (Sinha *et al.*, 2016; Bassis *et al.*, 2017). The profiling of the infant gut bacteria in the first year of life showed that this community is very dynamic and over time starts to mimic the adult gut bacterial profile. Previous studies that have investigated the adult gut bacterial profile have observed that this profile is distinctly different from the infant profile as it is generally stable, and has minimal fluctuations over time (Yatsunenko *et al.*, 2012; Rodríguez *et al.*, 2015). On the other hand, the gut bacterial profiles of infants exhibit high inter-individual diversity during the first few year of life (Yatsunenko *et al.*, 2012).

From this study, infants that were breastfed in the first year of life had significant differences in some bacterial taxa when compared to those who were not. There was no association between breast feeding and development of wheezing in this study. Formula fed infants in this study had increased proportions of *Bacteroidetes* and this has been shown in other studies (Penders *et al.*, 2006; Adlerberth and Wold, 2009). Other studies showed that exclusively breastfed infants had a higher composition of *Bifidobacteria,* and lower compositions of *E. coli* and *C. difficile* (Nambu, Shintaku and Ohta, 2004; Penders *et al.*, 2008). These studies found inconsistent data on the association between the mode of feeding and the development of wheezing illness in infants. A major limitation of this study was that the sequencing carried out was only able to characterize the bacteria up to genus level. This limited the level to which we were able to determine association between the infant stool bacteria and wheezing illness. Performing Whole Genome Shotgun sequencing on the Illumina platform would allow identification down to species level taxonomy (Ma, Prince and Aagaard, 2014).

The use of electricity as a primary fuel affected the infant bacterial composition but was not associated with wheezing in the first year of life. The taxa *Actinobacteria* and *Bifidobacterium* were significantly increased and *Proteobacteria* was significantly reduced in infant's stool whose households used electricity. We presume that the socioeconomic status which is represented by the use of electricity plays a role in influencing the gut bacterial composition. Previous studies by groups in different countries have shown that the socioeconomic status affects the gut bacterial composition (Tyakht *et al.*, 2013; Chong *et al.*, 2015; Mello *et al.*, 2016; Miller *et al.*, 2016). Studies in literature have also reported that the socioeconomic status affects the state of allergy, including wheezing and/or asthma in children (Forastiere *et al.*, 1997; Hafkamp-de Groen *et al.*, 2012; Galobardes *et al.*, 2015).

Most studies suggest that pet exposure has a protective effect on the development of allergic diseases including wheezing illness (Azad *et al.*, 2013; Nermes *et al.*, 2013; Tun *et al.*, 2017). However, we were unable to establish a relationship between pet exposure and the development of wheezing illness. Few studies from the systematic review investigated the direct or indirect involvement of pets in the development of wheezing illness (Murray *et al.*, 2005; Penders *et al.*, 2007, 2008; van Nimwegen *et al.*, 2011; Arrieta *et al.*, 2015; Stiemsma *et al.*, 2016). Only one of the studies included in the systematic review observed an association between pet ownership and development of wheeze and atopic sensitization (Murray *et al.*, 2005).

Previous studies investigating crowding have reported conflicting information on the association with the gut microbiota and wheezing (Infante-Rivard *et al.*, 2001; Rusconi *et al.*, 2005; Nicolaou *et al.*, 2008; Luijk *et al.*, 2015). In this study, there was no association between infants that experienced crowding either by having multiple siblings or by being in homes with large numbers of household members and the development of wheezing. However, differences were observed in the gut microbiota of infants who were exposed to crowding, versus those who were not. More studies investigating different crowding variables in association with development of wheezing in infancy would be beneficial in tackling this topic more exhaustively.

In this study, the living area had an effect on the infant stool bacteria but was not associated with the development of wheezing in infants. Infants living in rural areas had increased amounts of *Firmicutes*, when compared to those living in urban areas and townships. This is possibly due to the fact that diets higher in starch, consumed in rural communities tend to be associated with gut bacterial populations dominated by *Firmicutes* (Liefert, 2004; Flint *et al.*, 2012). Nevertheless other studies have found that children growing up in farm or rural areas

were less likely to experience wheezing (Riedler *et al.*, 2000; Von Ehrenstein *et al.*, 2000; Stein *et al.*, 2016).

Additionally, we were able to observe a relationship between the infant gut bacterial profile and wheezing illness, as shown by other studies (Murray *et al.*, 2005; van Nimwegen *et al.*, 2011; Arrieta *et al.*, 2015; Stiemsma *et al.*, 2016). From the bi-plots only one bacterial taxon (*Lactobacillales)* was significantly associated with wheezing, although we suspect that this was due to outliers and few data points for recurrent wheezers. Previous studies have shown that wheezing was associated with various bacterial taxa including *Bifidobacteria*, *Clostridia* and the FLVR (*Faecalibacterium, Lachnospira, Vellionella* and *Rothia)* bacteria (Verhulst *et al.*, 2008; Arrieta *et al.*, 2015; Stiemsma *et al.*, 2016). We assume that increasing the number of recurrent wheezers would show a stronger association to certain bacterial taxa.

This study had limitations which affected the data analyzed herein. One of these limitations was that the majority of infants did not have full data sets, they often had one or more missing samples, and this prevented the wheezing analysis from being carried out exhaustively. Another limitation was the incomplete clinical data profiles collected for the participants, which accounted for about 25%. This in turn hindered accurate analyses being carried out. Therefore it is important to have all the stool samples because other studies have observed that the first 100 days represent the critical period in which changes in the stool microbiota could be attributed to wheezing later on (Van Der Velden *et al.*, 2001; Arrieta *et al.*, 2015a; Stiemsma *et al.*, 2016).

Chapter 5 - General Conclusion

5.1. Conclusion

This study was able to profile the stool bacteria of the infants from birth to 12 months. In addition, we identified the risk factors that influenced the acquisition of specific bacterial groups or affected the stool diversity as a whole. This study was not able to conclusively determine any association between the infant wheezing and gut bacterial profile. We suspect that this was due to the inadequate number of recurrent wheezers within the first year of infant life.

We anticipate that future studies on recurrent wheezing in this cohort would require a much larger group of recurrent wheezers, with full data sets on risk factors, clinical data and all the required samples collected. A larger recurrent wheezing cohort would enable researchers to note if the four bacterial genera; *Faecalibacterium, Lachnospira, Vellionella* and *Rothia* also had a protective role in infants in South Africa (Arrieta *et al.*, 2015b; Stiemsma *et al.*, 2016). Further research will also be required to assess the association of *C. difficile* colonization on the increased risk of wheezing in infants. As found in Systematic Review, more research needs to be carried out to determine the influence of prebiotics, probiotics and breastfeeding in the development of wheezing in infants. Similarly, since recurrent wheezing is one of the precursors of asthma, there is a need for an additional follow-up of the wheezing infants in order to investigate the contribution of the early faecal microbiome in the development of asthma at school age.

5.2. References

Adlerberth, I. and Wold, A. E. (2009) 'Establishment of the gut microbiota in Western infants', *Acta Paediatrica, International Journal of Paediatrics*, 98(2), pp. 229–238. doi: 10.1111/j.1651-2227.2008.01060.x.

Arrieta, M. *et al.* (2015) 'Early infancy microbial and metabolic alterations affect risk of childhood asthma', *Science Translational Medicin*, 7(307).

Azad, M. B. *et al.* (2013) 'Infant gut microbiota and the hygiene hypothesis of allergic disease: impact of household pets and siblings on microbiota composition and diversity.', *Allergy, asthma, and clinical immunology : official journal of the Canadian Society of Allergy and Clinical Immunology*, 9, p. 15. doi: 10.1186/1710-1492-9-15.

Bassis, C. M. *et al.* (2017) 'Comparison of stool versus rectal swab samples and storage conditions on bacterial community profiles', *BMC Microbiology*, 17(1), pp. 1–7. doi: 10.1186/s12866-017-0983-9.

Chong, C. W. *et al.* (2015) 'Effect of ethnicity and socioeconomic variation to the gut microbiota composition among pre-adolescent in Malaysia', *Nature Publishing Group*, pp. 1–12. doi: 10.1038/srep13338.

Von Ehrenstein, O. S. *et al.* (2000) 'Reduced risk of hay fever and asthma among children of farmers', *Clinical and Experimental Allergy*, 30(2), pp. 187–193. doi: 10.1046/j.1365-2222.2000.00801.x.

Flint, H. J. *et al.* (2012) 'Microbial degradation of complex carbohydrates in the gut', *Gut Microbes*, 3, pp. 289–306.

Forastiere, F. *et al.* (1997) 'Socioeconomic status, number of siblings, and respiratory infections in early life as determinants of atopy in children', *Epidemiology*, pp. 566–570.

Galobardes, B. *et al.* (2015) 'Childhood wheezing, asthma, allergy, atopy, and lung function: different socioeconomic patterns for different phenotypes', *American Journal of Epidemiology*, 182(9), pp. 763–774. doi: 10.1093/aje/kwv045.

Hafkamp-de Groen, E. *et al.* (2012) 'The role of prenatal, perinatal and postnatal factors in the explanation of socioeconomic inequalities in preschool asthma symptoms: The generation R study', *Journal of Epidemiology and Community Health*, 66(11), pp. 1017–1024. doi: 10.1136/jech-2011-200333.

Infante-Rivard, C. *et al.* (2001) 'Family size, day-care attendance, and breastfeeding in relation to the incidence of childhood asthma', *American journal of epidemiology*, 153(7), pp. 653–8.

Liefert, W. (2004) 'Food security in Russia: economic growth and rising incomes are reducing insecurity', *Food Security Assess*, pp. 35–43.

Luijk, M. P. C. M. *et al.* (2015) 'Is parent-child bed-sharing a risk for wheezing and asthma in early childhood?', *European Respiratory Journal*, 45(3), pp. 661–669. doi: 10.1183/09031936.00041714.

Ma, J., Prince, A. and Aagaard, K. M. (2014) 'Use of whole genome shotgun metagenomics: A practical guide for the microbiome-minded physician scientist', *Seminars in Reproductive Medicine*, 32(1), pp. 5–13. doi: 10.1055/s-0033-1361817.

Mello, C. S. *et al.* (2016) 'Gut microbiota differences in children from distinct socioeconomic levels living in the same urban area in Brazil', *Journal of Pediatric Gastroenterology and*

Nutrition, 63(5), pp. 460–465. doi: 10.1097/MPG.0000000000001186.

Miller, G. E. *et al.* (2016) 'Lower neighborhood socioeconomic status associated with reduced diversity of the colonic microbiota in healthy adults', pp. 1–17. doi: 10.1371/journal.pone.0148952.

Murray, C. S. *et al.* (2005) 'Fecal microbiota in sensitized wheezy and non-sensitized non-wheezy children : a nested case – control study', pp. 741–745. doi: 10.1111/j.1365-2222.2005.02259.x.

Nambu, M., Shintaku, N. and Ohta, S. (2004) 'Intestinal microflora at 4 months of age and the development of allergy', *Allergology International*, 53(2), pp. 121–126. doi: 10.1111/j.1440-1592.2004.00315.x.

Nermes, M. *et al.* (2013) 'Perinatal pet exposure, faecal microbiota, and wheezy bronchitis: is there a connection?', *ISRN Allergy*, pp. 1–6. Available at: 10.1155/2013/827934.

Nicolaou, N. C. *et al.* (2008) 'Day-care attendance, position in sibship, and early childhood wheezing: A population-based birth cohort study', *Journal of Allergy and Clinical Immunology*, 122(3). doi: 10.1016/j.jaci.2008.06.033.

van Nimwegen, F. A. *et al.* (2011) 'Mode and place of delivery, gastrointestinal microbiota, and their influence on asthma and atopy.', *The Journal of allergy and clinical immunology*, 128(5), pp. 943–948. doi: 10.1016/j.jaci.2011.07.027.

Penders, J. *et al.* (2006) 'Factors influencing the composition of the intestinal microbiota in early infancy.', *Pediatrics*, 118(2), pp. 511–521. doi: 10.1542/peds.2005-2824.

Penders, J. *et al.* (2007) 'Gut microbiota composition and development of atopic manifestations in infancy: the KOALA Birth Cohort Study.', *Gut*, 56(5), pp. 661–667. Available at: 10.1136/gut.2006.100164.

Penders, J. *et al.* (2008) 'Toxigenic and non-toxigenic Clostridium difficile : determinants of intestinal colonisation and role in childhood atopic manifestations', *Gut*, 57(7), pp. 1025–1026. doi: 10.1136/gut.2007.143214.

Riedler, J. *et al.* (2000) 'Austrian children living on a farm have less hay fever, asthma and allergic sensitization', *Clinical and Experimental Allergy*, 30(2), pp. 194–200. doi: 10.1046/j.1365-2222.2000.00799.x.

Rodríguez, J. M. *et al.* (2015) 'The composition of the gut microbiota throughout life, with an emphasis on early life', *Microbial Ecology in Health & Disease*, 26(0), pp. 1–17. doi: 10.3402/mehd.v26.26050.

Rusconi, F. *et al.* (2005) 'Risk factors in the pre-, perinatal and early life (first year) for Wheezing in Young Children', *Ratio*, (2), pp. 47–51.

Sinha, R. *et al.* (2016) 'Collecting fecal samples for microbiome analyses in epidemiology studies', *Cancer Epidemiology, Biomarkers & Prevention*, 25(2), pp. 407–416. doi: 10.1158/1055-9965.EPI-15-0951.Collecting.

Stein, M. M. *et al.* (2016) 'Innate immunity and asthma risk in amish and hutterite farm children', *New England Journal of Medicine*, 375(5), pp. 411–421. doi: 10.1056/NEJMoa1508749.

Stiemsma, L. T. *et al.* (2016) 'Shifts in Lachnospira and Clostridium sp. in the 3-month stool microbiome are associated with preschool age asthma', *Clinical Science*, 130(23), pp. 2199–2207. doi: 10.1042/CS20160349.

Tun, H. M. *et al.* (2017) ‘Exposure to household furry pets influences the gut microbiota of infants at 3–4 months following various birth scenarios’, *Microbiome*, 5(1), p. 40. doi: 10.1186/s40168-017-0254-x.

Tyakht, A. V. *et al.* (2013) ‘Human gut microbiota community structures in urban and rural populations in Russia’, *Nature Communications*, 4, pp. 1–9. doi: 10.1038/ncomms3469.

Verhulst, S. L. *et al.* (2008) ‘A longitudinal analysis on the association between antibiotic use, intestinal microflora, and wheezing during the first year of life.’, *The Journal of asthma : official journal of the Association for the Care of Asthma*, 45, pp. 828–832. doi: 10.1080/02770900802339734.

Yatsunenko, T. *et al.* (2012) ‘Human gut microbiome viewed across age and geography’, *Nature*, 486, pp. 222–227. doi: 10.1038/nature11053.